Genomic and flow cytometric studies of *Clostridium sporogenes*, a non-toxigenic surrogate for *Clostridium botulinum*

By

Mark Bradbury

Submitted in fulfilment of the requirements for the Degree of

Doctor of Philosophy

School of Health Sciences

Federation University

PO Box 663
University Drive, Mount Helen
Ballarat, Victoria 3353,
Australia

June 2014

Statement of authorship

Except where explicit reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis by which I have qualified for or been awarded another degree or diploma. No other person's work has been relied upon or used without due acknowledgment in the main text and bibliography of the thesis.

Mark Bradbury
June, 2014

Acknowledgments

Foremost, I would like to express my deep and sincere gratitude to my supervisors Dr. Janelle Brown and Prof. Frank Vriesekoop for their support, challenging questions, critiques, patience and trust throughout this thesis.

To the army of associate supervisors (formal or otherwise): Dr. Peter Aldred, Dr. David Bean, Dr. Belinda Chapman I would like to express my thanks for providing support, inspiration and your time. I also thank Federation University and CSIRO for providing a scholarship to support my studies.

To members of the CSIRO Food Microbiology team, past and present, I would like to express my thanks for your motivation, insights and friendship over the journey.

I would like to acknowledge Debra Birch and Nicole Vella from the Microscopy Unit of Biological Sciences at Macquarie University for their assistance with the TEM sample preparation.

Finally I would like to thank my family for their support, in particular my parents who have provided encouragement throughout my life and inspiration to continue to learn. And above all to my wife, Emily, for her many sacrificed weekends, encouragement and understanding over the last years.

List of presentations and papers

Published Manuscripts

- **Bradbury, M.**, Greenfield, P., Midgley, D., Li, D., Tran-Dinh, N., Vriesekoop, F., & Brown, J. L. (2012). Draft genome sequence of *Clostridium sporogenes* PA 3679, the common non-toxigenic surrogate for proteolytic *Clostridium botulinum*. *Journal of bacteriology*, *194*(6), 1631-1632.

Selected Conference Posters

- **Bradbury, M.;** Vriesekoop, F., Brown, J.L. (2013). *Comparative genomics of Clostridium sporogenes with reference to Group I Clostridium botulinum.* 8th International Conference on the Molecular Biology and Pathogenesis of the Clostridia (CLOSTPATH 8), Cairns, QLD, 22 - 26 October 2013.

- **Bradbury, M.;** Greenfield, P.; Chapman, B.; Tran-Dinh, N.; Midgley, D.; Li, D.; Vriesekoop, F., Brown, J.L..(2012). *Evaluation of Clostridium sporogenes PA 3679 as a non-toxigenic surrogate and model for studying the thermal inactivation of spores of proteolytic Clostridium botulinum.* 5th European Spores Conference, London, April 16-19, 2012.

- Brown, J.L.; Turnbull, L.; Stone, G.; Bull, M**.; Bradbury, M.;** Tran-Dinh, N.; Whitchurch, C.; Chapman, B. *Single cell resolution of physiological states in Clostridium sporogenes spore populations following exposure to moist heat.* 5th European Spores Conference, London, April 16-19, 2012.

Table of Contents

List of Figures

List of Tables

Summary

Clostridium sporogenes and Group I *Clostridium botulinum* are two bacterial species belonging to the same phylogenetic group, primarily differentiated by the expression of botulinum neurotoxin. Both organisms are of significant commercial importance in regards to the spoilage of and/or disease in thermally processed food products due to their ability to form heat resistant spores. As such, these species are often used as the target organism for the design of thermal inactivation processes, particularly in regards to thermal sterilisation. Two specific aspects pertaining to these organisms were investigated to further enhance knowledge with respect to their use in thermal processing studies: the genetic relationship between *C. sporogenes* and Group I *C. botulinum*; and the mechanism of heat resistance in spores of *C. sporogenes.*

The first part of this thesis describes the assembly, annotation and analysis of the draft genome of *C. sporogenes* PA 3679 (the most widely used surrogate for Group I *C. botulinum*). These data allowed identification of unique variants genes related to spore germination, analogous toxin regions and mobile elements between species. MLST analysis revealed that phylogeny was an ineffective indicator of toxigenicity in this group and thus prompted a pan-genomic analysis. The pan-genome of *C. sporogenes/*Group I *C. botulinum* was found to consist of 8799 coding sequences (CDS's) and a core genome consisting of 1590 CDS's. Analysis of this pan-genome revealed the significant role that mobile genetic elements have played in genetic diversity within this group of organisms.

The second part of this thesis investigated the heat inactivation of *C. sporogenes* PA 3679 spores in regards to structural variation and population heterogeneity. A novel flow cytometric approach was developed and utilised to investigate isothermal spore inactivation; and implications of the impact of NaCl on the intrinsic variability throughout this process and during a subsequent recovery period. Based on these approaches a possible mechanistic description for the thermal inactivation of spores was developed.

Together, these studies present significant evidence supporting the continued suitability of *C. sporogenes* as a surrogate for Group I *C. botulinum*, whilst also enhancing the understanding of clostridial spore inactivation during a moist heat process.

Chapter 1 – Introduction

Thermal processes such as pasteurisation are amongst the most widely used methods by manufacturers to extend the shelf life of foods and ensure microbiological safety. The application of heat, either through sterilisation, pasteurisation, or simply cooking inactivates micro-organisms potentially harmful to humans or responsible for spoilage. Certain species of bacteria form environmental resistant, dormant spores which are capable of surviving temperatures that would be lethal to vegetative cells. These spore forming bacteria present the greatest microbial hazard for manufacturers producing products that have been thermally processed.

The bacterial pathogen of greatest concern in thermally processed foods designed for extended shelf life is *Clostridium botulinum*. This species; is able to form spores capable of surviving the heating process, is an anaerobic organism capable of outgrowth under oxygen depleted conditions as found in many sealed products; and produces the most lethal natural toxin known. As such, extensive studies have been conducted to ensure that the heating process controls and product criteria provide an appropriate level of assurance against the outgrowth of spores.

Increasingly, studies have employed surrogate organisms to use instead of *C. botulinum*. The use of a surrogate for process verification and thermal inactivation has become an increasingly important tool. The impetus for this is both safety concerns associated with the toxin, as well as increasingly strict regulatory controls concerning the handling and possession of agents classified as potential bioterrorism threats.

The *Clostridium sporogenes* strain putrefactive anaerobe (PA) 3679 is a widely used surrogate for Group I *C. botulinum.* It is indistinguishable from Group I *C. botulinum* on the basis of morphology, biochemistry and growth characteristics. However, two specific characteristics differentiate the strain, firstly, the strain does not produce neurotoxin and secondly, spores of this strain have been shown to exhibit significantly higher heat resistance properties. These two properties identify this strain as an appropriate surrogate for Group I *C. botulinum* in the verification of thermal processes.

Despite its widespread use as a surrogate organism, *C. sporogenes* PA 3679, when compared to a number of other important commercially significant strains, remains poorly characterised. Whether the strain is a *C. botulinum* that has lost the ability to form neurotoxin or is an evolutionary distinct organism is unknown. Furthermore, potential contributors to the comparatively high heat resistance of its spores have not been identified.

Given the specific characteristics of *C. sporogenes* PA 3679 further characterisation presents an opportunity to provide insights into the heat resistance properties of spores and the evolution of toxigenicity in *C. botulinum*. Furthermore, a greater understanding of how this strain relates to Group I *C. botulinum* may inform on the appropriateness of using the organism as a surrogate.The aim of this research (as shown in Figure 1.1) was to provide improved understanding of the suitability of *C. sporogenes* as a surrogate for Group I *C. botulinum*.

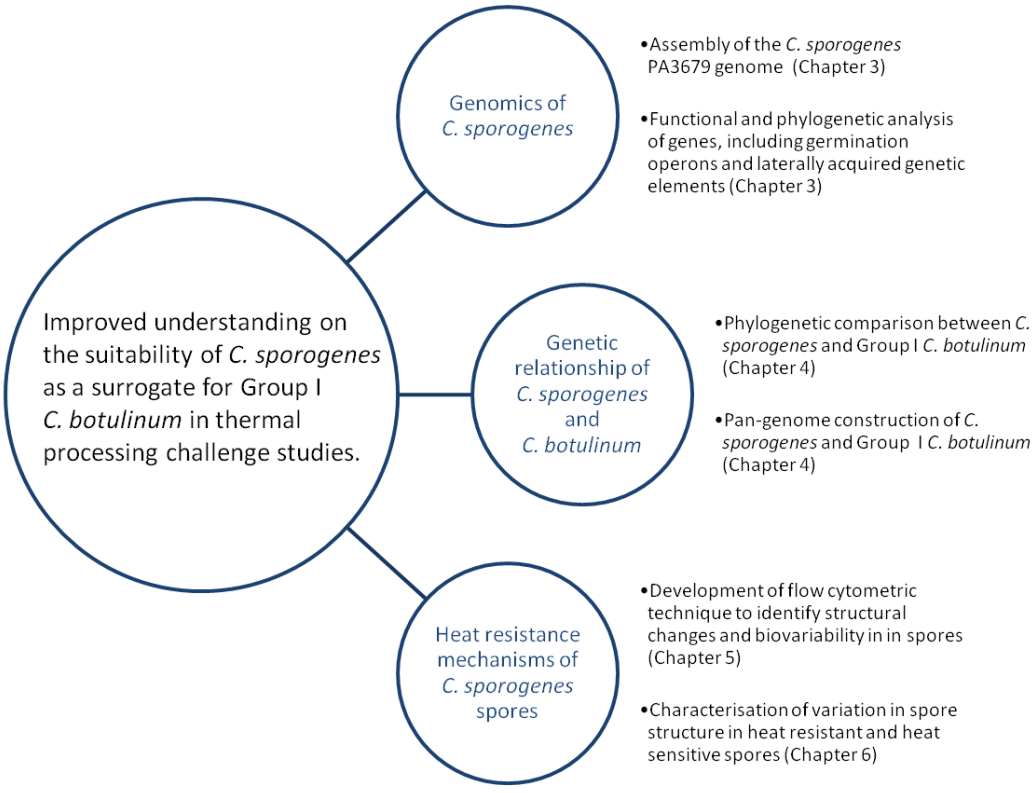

Figure 1.1 Graphical representation of the objectives of this study.

The objectives of this study were:

1) The assembly of a draft whole genome of *C. sporogenes* PA 3679
2) To utilise this genome to conduct a comparison of selected genetic locations including analogous toxin encoding regions and germination operons to Group I *C. botulinum*
3) To identify the phylogenetic relationships of *C. sporogenes* using whole genome approaches
4) To develop and validate a flow cytometric method to investigate the heterogeneity of spores during a thermal process
5) To investigate variability of heat resistance properties of *C. sporogenes* and associate these characteristics with the structural conformation of spores.

Chapter 2 - Literature Review

Clostridium botulinum and *Clostridium sporogenes* are amongst the most important bacteria that influence the design of processes aimed at ensuring the safety of thermally treated low-acid (pH ≥ 4.6) foods. Group I *C. botulinum* is the most heat resistant pathogen of concern in these products, while *C. sporogenes* is widely employed as a non-toxigenic surrogate organism for *C. botulinum* and also as a process target organism (NACMCF, 2010). The relationship between these two species is poorly understood and confusing, with the only unambiguous differentiation between species the production of botulinum neurotoxin (BoNT), the causative agent of botulism.

As a background to the objectives described in the introduction, this review will consider aspects of the taxonomy and genetic relationship between *C. botulinum* and *C. sporogenes*. In addition it will also include a discussion of selected factors contributing to the survival of spores under thermal stress, in particular spore structural complexity and population biovariability.

2.1 *Clostridium botulinum*

Whilst classified as a single bacterial species, *C. botulinum* comprises a phylogenetically and physiologically diverse collection of anaerobic, Gram-positive endospore forming rods unified by their ability to form BoNT.

The original isolation of *C. botulinum* was by van Ermengem (1897) from a botulism outbreak associated with smoked ham in Belgium. As subsequent botulism outbreaks were investigated it was established that the isolates exhibited varying physiological characteristics and toxin-related serological properties (Erbguth, 2004). Despite this high degree of physiological diversity all organisms that produced BoNT were designated as the species *C. botulinum* (Prévot, 1953); a practice that has persisted notwithstanding later observations of significant genetic diversity within the species. However, *C. botulinum* definition is based on BoNT production has since become less

definitive with the isolation and subsequent naming of *Clostridium baratti* (Hall et al., 1985) and *Clostridium butyricum* (McCroskey et al, 1986) that both produce BoNT and the reclassification of some *C. botulinum* into the species *C. argentinense* (Suen et al., 1988).

Classification of *C. botulinum* isolates is generally made via two systems, serotyping of BoNT and/or division into one of four physiologically distinct groups. Serotyping of isolates is achieved via neutralization of expressed BoNT with one of eight serologically distinct antitoxins. Prior to 2013 there were only seven BoNT serotypes identified (A-G) (Smith et al., 2005), however recently an eighth toxin type designated type H has been identified (Barash and Arnon, 2013). Within this alphabetic serotyping scheme are further subtypes based on amino acid sequence variation and designated numerically (e.g. A1-5, B1-3) (Smith et al., 2005). Bivalent strains of *C. botulinum* capable of expression of more than just one BoNT type have also been identified and are designated by capitalisation of the predominant toxin with the secondary toxin following in lower case (e.g. Ab, Ba, Bf).

In addition to serotyping, *C. botulinum* isolates may be divided into four physiological groups (Groups I-IV). It has been noted that the significant differences in physiology and DNA homology would justify division into four separate species; however the medical significance of botulinum toxin has led them to be retained as one species (Sneath, 1986; Collins and East, 1998). Isolates within each group may express differing toxin serotypes and conversely isolates in different groups may exhibit the same toxin type, revealing discordant phylogenies between toxin type and host bacteria (Hill and Smith, 2013). A summary of these characteristic of the groups is presented in Table 2.1. Group I (proteolytic) *C. botulinum* are mesophilic sporeformers that digest complex proteins, these strains produce the most heat resistant spores and may express BoNT serotypes A, B, F and/or H. Group II (non-proteolytic) *C. botulinum* strains produce serotypes B, E and F. Group III *C. botulinum* produce BoNT serotypes C and D, whilst Group IV *C. botulinum* produce BoNT serotype G.

Within each group both toxigenic and non-toxigenic isolates have been reported, however as the defining feature of *C. botulinum* is toxin production such strains are often taxonomically separated as different species, with the exception of Group II.

Such species associated with individual groups include *C. sporogenes* (Group I), *Clostridium novyi* and *Clostridium haemolyticum* (Group III) and *Clostridium subterminale* (Group IV).

Table 2.1 Characteristics of *C. botulinum* groups (Adapted: Peck, 2009)

Neurotoxigenic clostridia	***C. botulinum Group I (proteolytic C. botulinum)***	***C. botulinum Group II (non-proteolytic C. botulinum)***	***C. botulinum Group III***	***C. botulinum Group IV (C. argentinense)***
Neurotoxins formed	A, B, F, H	B, E, F	C, D	G
Non-neurotoxigenic equivalent clostridia	*C. sporogenes*	No species name given	*C. novyi*	*C. subterminale*
Optimum growth temperature	37°C	25°C	40°C	37°C
Minimum growth temperature	10-12°C	2.5-3.0°C	15°C	-
Minimum growth pH	4.6	5.0	5.1	-
NaCl concentrations preventing growth	10%	5%	-	-
Spore heat resistance in phosphate buffer[a]	$D_{121°C}$ = 12.6 seconds	$D_{121°C}$ = 0.02 seconds	$D_{121°C}$ = 1.05 seconds	$D_{121°C}$ = 1.29 seconds

[a]$D_{121.1°C}$ values were calculated assuming a z-value of 10.

2.2 Botulism

The medical significance of *C. botulinum* relates to BoNT, the causative agent of human and animal botulism (Johnson and Montecucco, 2008). These toxins are among the most potent natural toxins known (Peck, 2009), with as little as 30 ng being potentially fatal (Johnson and Bradshaw, 2001). There are three major types of human-related botulism, foodborne botulism, infant/intestinal (adult) botulism and wound botulism. Foodborne botulism refers to the ingestion of pre-formed toxin produced in foods. Infant/intestinal botulism occurs due to ingestion of *C. botulinum* spores into gastrointestinal tracts of people with underdeveloped intestinal microbiota which are typically capable of preventing the colonisation of *C. botulinum* (Koepke et al., 2008). Currently wound botulism is most commonly associated with drug use, particularly heroin (Yuan et al., 2011; MacDonald et al., 2013).

Botulism in humans occurs when BoNT interrupts the transmission of nerve impulses to the muscles. In the early 1900's, 71% of botulism cases in the US were fatal but by 1970's this had decreased to 15% (Hauschild and Dodds, 1993). Presently with the increased availability of antitoxin and supportive therapy the fatality rate has dropped to approximately 5-10% of cases (Peck, 2006). The onset of symptoms usually occurs within 12-36 h but can occur within 2 hours or take as long as 8 days (McLauchlin et al., 2006). The BoNT toxin initially causes paralysis to neuromuscular junctions of the head and neck producing symptoms including double vision, drooping eyelids, thickness of speech, and inability to swallow or speak clearly. If left untreated death results from airway obstruction or respiratory failure (McLauchlin et al., 2006).

2.2.1 Foodborne botulism associated with Group I *C. botulinum*

Foodborne botulism is relatively rare in developed countries, particularly in Australia where between 1942-89 there have been only five recorded outbreaks with a total of fifty three people affected and nine fatalities (Hauschild, 1993). Four of these outbreaks were linked to vegetables, while one outbreak was associated with canned tuna. More recently in 2007 a male in Victoria was exposed to type A toxin in a microwave meal of pre-packaged nachos (FSANZ, 2007) resulting in a prolonged hospitalisation from which he recovered. Foodborne botulism is still more prevalent in some geographical areas, particularly regions with a strong tradition of home canning. In Europe in 1999/2000 over 2,500 cases of foodborne botulism were reported, primarily in Russia, which were predominantly attributed to smoked, salted and fried fish, and home canned vegetables (Peck, 2006; Lindström et al., 2009b).

Despite the relative rarity of botulism the severity of the illness and prolonged treatment required by patients make the potential cost high. Setlow and Johnson (1997) predicted that in the US each case of botulism could cost approximately $30 million.

Strains of *C. botulinum* residing in Groups I and II are most commonly implicated in foodborne botulism. The physiological difference between these groups is reflected in the types of food commodities in which they are most commonly found to present a risk. Spores of Group I (proteolytic) *C. botulinum* exhibit the highest heat resistance and thus are most commonly associated with thermally processed shelf-stable low

acid, hermetically sealed foods. Group II (non-proteolytic) *C. botulinum* are typically related to foods which have been mildly heat processed and refrigerated and/or marine foods (Lindström et al., 2006) due to their significantly lower heat resistant and optimum growth temperature compared to their Group I counterparts. Members of Group III *C. botulinum* are typically associated with cases of animal botulism, whilst Group IV *C. botulinum* isolates are rarely associated with botulism cases.

Historically the prevention of foodborne botulism has relied on severe thermal processing, the basis of which relates to the inactivation of Group I *C. botulinum* spores. The intrinsic properties of the food product determine the level of treatment required. Foods intended for ambient storage are classified based on their pH as either "low acid" (pH ≥4.6) or "acid foods" (pH<4.6) (Moir et al., 2001). Since commercial processes have become more regulated the incidences of outbreaks in commercially produced food products has become less common and currently most outbreaks are attributed to home canning (Peck, 2009; Pflug, 2010).

2.3 Molecular aspects of BoNT in Group I *C. botulinum*

The botulinum protein toxin complex comprises both BoNT proteins and several associated non-toxic proteins (ANTP's) including hemagglutinin (HA) and non-toxic non-hemagglutinin (NTNH) proteins (Gu and Jin, 2013). The ANTP components of this protein complex have been shown to enhance the potency, increase the stability of the toxin against biophysical processes (Chen et al., 1998) and facilitate transfer of the toxin through the gastro-intestinal tract (Fujinaga et al., 2013).

2.3.1 Genetic organisation of BoNT genes in Group I *C. botulinum*

The genetic organisation of the toxin region is well described across a variety of serotypes and strains (Figure 2.1). The coding sequences of the BoNT and ANTP's are typically co-located with the transcriptional regulator, *botR* in a single cluster (Raffestin et al., 2005). Overexpression of *botR* has been shown to induce the production of both BoNT and ANTP's (Marvaud et al., 1998).

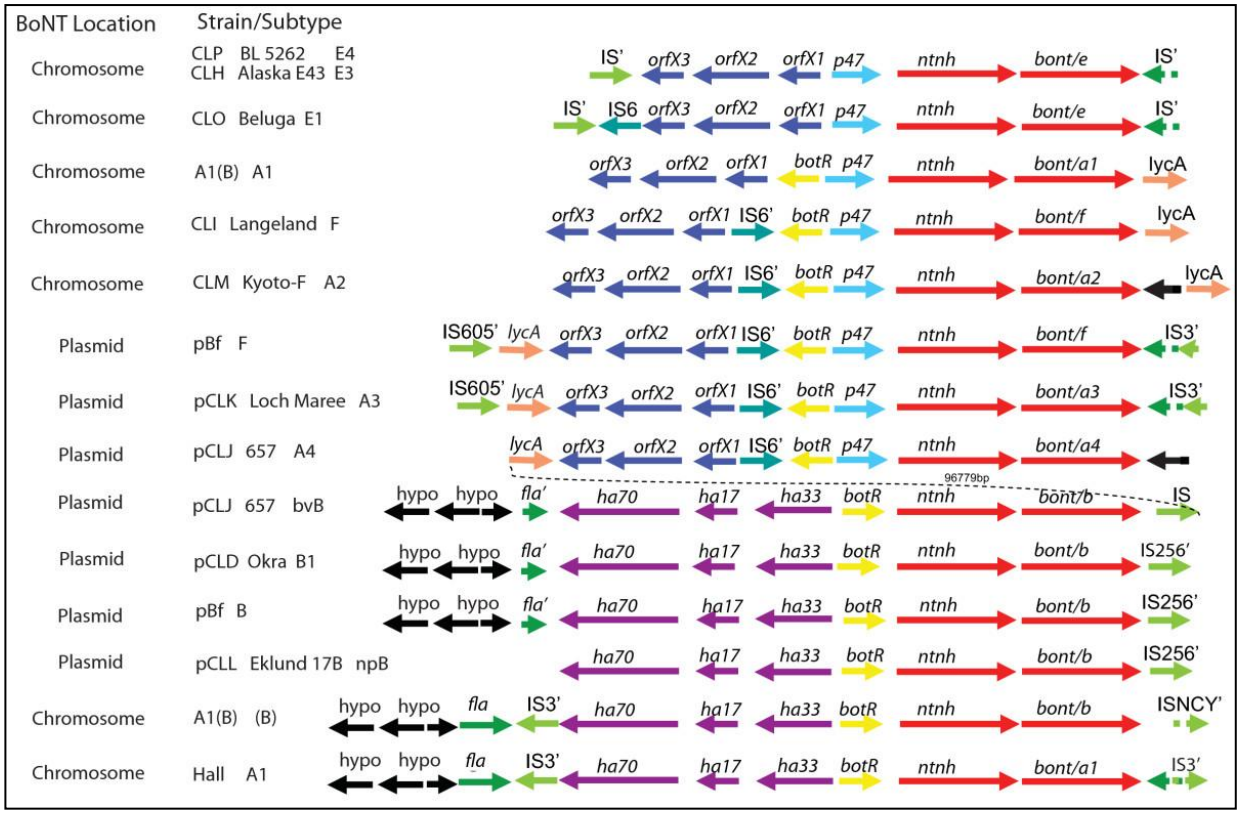

Figure 2.1 Genetic organisation of BoNT genes and location on sequenced strains of *C. botulinum* (Source: Hill et al., 2009)

Two conserved toxin gene cluster arrangements have been identified in Group I *C. botulinum* strains, the *ha*$^+$ complex (*ha70-ha17-ha33-botR-ntnh-bont)* and the *orfX*$^+$ complex (*orfX3-orfX2-orfX1-botR-p47-ntnh-bont*) (Hill et al., 2009). The *ha*$^+$ complex is associated with A1 and B BoNT serotypes and contains three hemagglutinins of varying molecular weight. These hemagglutinins proteins are associated with disruption of the intestinal epithelium permitting passive diffusion of the toxin (Matsumura et al., 2008).

The *orfX*$^+$ complex is associated with types A2, A3, A4 and F serotypes in Group I *C. botulinum* and also a number of A1 in Ab seroytpes. This complex lacks the three hemagglutinins of the *ha*$^+$ complex, however contains a group of three reading frames (*orfX*, *orfX2* and *orfX3*) encoding proteins of unknown function.

2.3.2 Genomic location of BoNT genes

Group I *C. botulinum* have been shown to contain genes encoding *BoNT/A, BoNT/B, BoNT/F* and most recently *BoNT/H*. These genes may be located on the chromosome and/or be plasmid-borne. The location of these genes does not appear to be serotype specific.

The location of the chromosomally bound *BoNT* genes within Group I strains appears conserved (Hill et al., 2009) either within the *arsC* operon or the *oppA/brnQ* operon. The *arsC* operon is located adjacent to an arsenate reductase gene cluster; it has been shown to contain *BoNT/A1*, *BoNT/A2* and *BoNT/F*. The *oppA/brnQ* operon is located downstream of the *arsC* operon and has been shown to contain the *BoNT/A1* gene, as well as the *BoNT/A5* and *BoNT/B* genes.

In addition to being located on the chromosome of Group I *C. botulinum BoNT/B* occur widely on plasmids (Smith et al., 2007; Franciosa et al., 2009; Umeda et al., 2012). Strains harbouring *BoNT/B* genes on plasmids have been shown to lose their toxigenicity after serial passages (Umeda et al., 2012). In a study by Franciosa et al., (2009), the authors found that *BoNT/B1* were typically located on plasmids while the *BoNT/B2* were located on the chromosome. This is in contrast to Umeda et al., (2013) who studied the location of *BoNT/B* in fifteen strains and indicated that the genomic location was not correlated with BoNT/B subtype. The variation between these results may be associated with the geographically source of the isolates.

In non-Group I strains, BoNT genes may be found in numerous other locations indicative of the mobility of the toxin cluster. In Group II *C. botulinum* BoNT genes are located both chromosomally in the *rarA* operon and on plasmids. In Group IV *C. botulinum* strains *BoNT/G* are located on plasmids (Zhou et al., 1995). Group III *C. botulinum* have *BoNT/C* and *BoNT/D* encoded on bacteriophages; as such the stability of toxigenicity in these strains is low. However, strains cured of these bacteriophages are capable of reconversion to toxigenicity by the addition of phage particles (Inoue and Iida, 1970; Eklund et al., 1972).

2.4 *Clostridium sporogenes*

The non-neurotoxigenic species *C. sporogenes* was first isolated by Metchnikoff (1908) from the faeces of healthy individuals. Differentiation between *C. sporogenes* and Group I *C. botulinum* is solely on the basis of toxigenicity. The International Judicial Commission of the International Committee on Systematic Bacteriology approved the opinion that "…*Clostridium botulinum* is conserved for toxigenic strains and *Clostridium sporogenes* is conserved for non-toxigenic strains" (JCICSB, 1999). Based on morphological and growth characteristics *C. sporogenes* is indistinguishable from Group I *C. botulinum* without detection of BoNT production. Genetically, Group I *C. botulinum* and *C. sporogenes* exhibit > 99.7% 16S *rRNA* homology and form a single phylogenetic unit (Hutson et al., 1993), indicative of the high degree of genetic relatedness of these two species. In contrast, Group I and Group II *C. botulinum* exhibit only ~90-91% 16S *rRNA* homology (Hutson et al., 1993); this value is significantly below the arbitrary "rule of thumb" species cut-off of 97% homology. (Stackebrant and Goebel, 1994).

C. sporogenes is a prominent organism in investigations into the thermal sterilisation of low-acid products. The organism is important with respect to its role as a spoilage organism and also as a surrogate organism for *C. botulinum* in the derivation of thermal process schedules. Despite the basis for its association as a surrogate based on its thermal resistance it has also been employed in studies of alternative non-thermal novel food sterilization processes such as high pressure processing (Mills et al., 1998; Bull et al., 2009). Further, it has been employed as a non-pathogenic research model organism to inform on the response of *C. botulinum* in physiological studies and transcriptional studies (Bassi et al., 2013; Taylor et al., 2013).

A particularly important strain of *C. sporogenes* is designated as PA 3679 (ATCC 7955, NCTC 8594). This strain has significantly higher spore heat resistance compared to group I *C. botulinum* (Brown et al., 2012). The isolate was originally isolated by E. J. Cameron (as cited in Townsend et al., 1938) from spoiled canned corn. Despite initially being taxonomically placed within a separate group of organisms referred to as "Putrefactive Anaerobes" (PA) it is now widely considered a strain of *C. sporogenes;* although there is no clear record of when the species classification was assigned. The reported D_{121}-values of this strain are as high as 1.5 minutes which is significantly higher than the reported heat resistance value of Group

I *C. botulinum* (D_{121}=0.21 minutes) indicative of a significantly higher heat resistance than any Group I *C. botulinum* strain (Figure 2.2). Due to this significantly higher heat resistance this strain is described as an "excellent surrogate for *C. botulinum* when used in inoculated pack studies to validate thermal processes for low-acid canned foods" by the National Advisory Committee on Microbiological Criteria for Foods (NACMCF, 2010). No detailed genetic characterisation of this strain has been conducted and hence the exact relationship of this strain to Group I *C. botulinum* remains poorly understood.

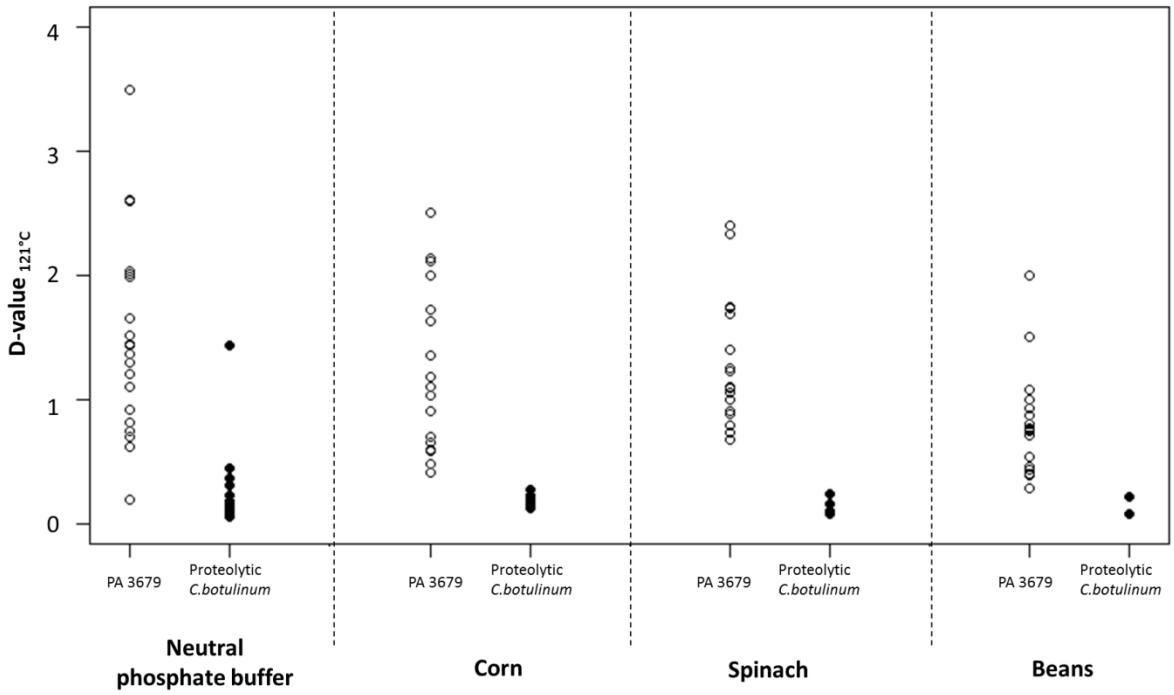

Figure 2.2 Comparison of the derived D121°C values of *C. sporogenes* PA 3679 and Group I *C. botulinum* compiled from a range of studies (n= 32). For references refer to Brown et al., 2012.

2.5 Molecular typing of Group I *C. botulinum* and *C. sporogenes*

While *C. botulinum* can broadly be described in terms of either Group I-IV or serotyping (A-H) based on the organisms' physiological characteristics or BoNT type respectively, neither method provides suitable discriminatory power for individual strains. As such various alternative molecular subtyping approaches have been used to differentiate strains of Group I *C. botulinum* including pulsed field gel electrophoresis

(PFGE) (Lin and Johnson, 1995; Franciosa et al., 2004; Nevas et al.,2005), randomly amplified polymorphic DNA analysis (RAPD) (Hyytiä et al., 1999; Franciosa et al., 2004; Pourshafie et al., 2005), multi-locus variable-number tandem-repeat analysis (MLVA) (Macdonald et al., 2008), multi-locus sequence typing (MLST) (Jacobson et al., 2008; Lúquez et al., 2012, Olsen et al., 2014), amplified fragment length polymorphism DNA analysis (AFLP) (Keto-Timonen et al., 2005), riboprinting (Skinner et al., 2000) and repetitive-element sequence-based PCR (Hyytiä et al., 1999). Such methods have all been shown to offer rapid and suitable levels of discrimination between strains for clinical and epidemiological studies. However, the majority of these studies have been applied to subtyping strains for diagnostic purposes rather than inference of the phylogenetic relationship between strains.

Whilst predominately focussing on discrimination within Group I *C. botulinum* typing studies often have included small numbers of *C. sporogenes* amongst the isolates. MacDonald et al., (2008) conducted a study to investigate the applicability of variable number tandem repeat (VNTR) markers to discriminate between seventy three *C. botulinum* strains and two *C. sporogenes*; of the ten VNTR loci used, six of these were amplified within both *C. sporogenes* strains indicating that the method was unable to unambiguously separate the two species. These results were similar to the findings of Pourshafie et al., (2005) who employed ribosomal DNA restriction analysis to investigate twelve Type A *C. botulinum* isolates and three environmental *C. sporogenes* strains. Whilst able to distinguish between the *C. botulinum* and *C. sporogenes* strains tested, a large of the resultant fragments from *C. sporogenes* were shared by six *C. botulinum* strains.

A novel MLST study consisting of partial sequencing of six genes, *adk, atoH, gyr,B, proC, rpoD* and *spo0A* was recently investigated with forty five *C. botulinum* strains and six *C. sporogenes* strains (Olsen et al., 2014). The study found clustering of five *C. sporogenes* isolates with a single outlier *C. sporogenes* isolate which was more closely related to *C. botulinum*. Similar results were observed in a study by Jacobson et al., (2008) who developed a MLST scheme using *mdh, aceK, rpoB, aroE, hsp60, oppB* and *recA.* In this study seventy three *C. botulinum* strains were used with only a single *C. sporogenes* isolate which clustered in the same clade as the majority of *C. botulinum* strains tested. The results indicated that similar levels of divergence can be observed between individual strains of *C. sporogenes* and Group I *C. botulinum*.

Comparative genomic hybridisation (CGH) studies have been employed by numerous authors to provide large scale investigations into multiple strains of *C. botulinum.* In particular a microarray based on the genome of *C. botulinum* ATCC 3502 has been widely employed. This microarray has been utilised for CGH studies of the diversity of Group I *C. botulinum;* including sixty strains of Type A (Carter et al., 2009) and thirty two Type B isolates (Lindström et al., 2009a). Significant variation in the neurotoxin sequence, gene content and genome arrangement was observed between the ATCC 3502, 19397, Hall, Allergan Hall and UMASS strains of A1 Group I *C. botulinum* (Fang et al., 2010).

Other authors have developed focused DNA microarrays (Raphael, 2010; Vanhomwegen et al., 2013) using variable regions however both of these rely to a large degree on BoNT related genes. As such, they present a useful tool for describing whether an isolate is *C. sporogenes* or *C. botulinum,* however these tools have limited applicability towards elucidating the evolutionary backgrounds of isolates that do not form toxins.

These molecular typing approaches highlight the difficulties associated with the differentiation between Group I *C. botulinum* and *C. sporogenes*. They also indicate a strong bias towards toxigenic strains over non-toxigenic strains.

2.6 Whole genome sequencing studies of Group I *C. botulinum* and *C. sporogenes*

As of December 2013 there are ten completed Group I *C. botulinum* genomes located on the NCBI reference sequence genome database (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/). In addition a number of draft genomes also are present in the NCBI database. Relative to other foodborne pathogens this is a small number, however with the advent of next-generation sequencing it is likely that this number will grow exponentially over the coming years.

The first completed whole genome sequence for a *C. botulinum* was that of a Group I Type A strain (ATCC 3502) (Sebaihia et al., 2007). This genome consists of a 3.9 Mbp chromosome and a 16 kbp plasmid, with 3650 and 19 predicted genes, respectively (Sebaihia et al., 2007). This genome has been widely employed as the

reference genome in subsequent Group I *C. botulinum* sequencing studies. Furthermore, this genome was used as the basis for the construction of the Group I *C. botulinum* microarray utilised for CGH studies as discussed in the previous section of this review. The current listing of completed Group I *C. botulinum* genomes present in the NCBI genome database are presented in Table 2.2.

In contrast to a high divergence shown across the genomes of all *C. botulinum*, a high degree of similarity has been observed across all Group I *C. botulinum* genomes. These studies coupled with CGH studies have managed to identify some key regions of divergence including resistance to arsenic and cadmium (Lindström et al., 2009a); and variation in the flagellar glycosylation island (Sebaihia et al., 2007).

Table 2.2 List of completed Group I *C. botulinum* genomes available from the NCBI reference database (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/)

Genome	Genome size* (Mbp)	Toxin serotype (Genomic location)	NCBI Accession no.
***C. botulinum* Kyoto**	4.16	A2 (plasmid)	NC_012563
***C. botulinum* Loch Maree**	3.99	A3 (plasmid)	NC_010520
***C. botulinum* ATCC 19397**	3.86	A1 (chromosome)	NC_009697
***C. botulinum* ATCC 3502**	3.89	A1 (chromosome)	NC_009495
***C. botulinum* Hall**	3.76	A1 (chromosome)	NC_009698
***C. botulinum* Okra**	3.96	B1 (plasmid)	NC_010516
***C. botulinum* Ba4 657**	3.98	Ba (plasmid)	NC_012658
***C. botulinum* Langeland**	4.00	F (chromosome)	NC_009699
***C. botulinum* 230613**	3.99	F (chromosome)	NC_017297
***C. botulinum* H04402 065**	3.92	A5 (chromosome)	NC_017299

Furthermore these studies have provided significant insights into the occurrence of silent BoNT genes such as a truncated BoNT/B3 gene in the Type A5 Strain H04402 (Carter et al., 2010). Further diversity of BoNT gene location was observed in sequencing *C. botulinum* strain Af84 (Dover et al., 2013). They found that this strain contained the three toxin genes, BoNT/A2, BoNT/F4 and BoNT/F5. The BoNT/A2 and BoNT/F5 were co-located on the chromosome of the strain whilst the BoNT/F5

was located on a 246 kb plasmid. This was the first strain to contain two variants of the same BoNT serotype (F) and have both plasmid and chromosomally encoded BoNT genes.

Whole genome sequencing studies have also provided significant insights into non-Group I *C.* botulinum. A number of Group III *C. botulinum* genomes have also been released (Skarin et al., 2011; Hassan et al., 2013). These genomes were shown to have a large diversity of plasmids. The chromosomes of these strains are relatively small (<3 Mbp) indicating a low tolerance for redundant genetic material (Skarin et al., 2011). In the case of the BoNT C/D BKT015925, a total of 13.5% of the genome was constituted by five large plasmids indicating a highly fluid genome in this strain.

2.7 Properties of spores of *Clostridium spp.*

2.7.1 Sporulation

A number of species of bacteria, most notably members of the genera *Bacillus* and *Clostridium*, are capable of forming endospores in response to nutrient limitation. Dormant spores may survive for extremely long periods under environmental conditions otherwise lethal to their vegetative counterparts. Upon sensing a favourable environment the spore will germinate and may return to active growth.

Much of the fundamental understanding of spores has arisen from studies on the model organism *Bacillus subtilis* rather than *Clostridium spp*. However, the large majority of sporulation genes appear highly conserved across species, and the general structure and physiology of spores is tacitly considered similar between *Bacillus spp.* and *Clostridium spp.* (Stragier, 2002).

The spore life cycle is a complex cascade of different stages. There are numerous reviews on the process of sporulation in *Bacillus* spp. as an excellent model for cellular differentiation (Errington, 1993; Errington, 2003). Sporulation is primarily instigated due to nutrient depletion or high cell density (Grossman, 1995), although other factors may induce sporulation including the Krebs cycle status, environmental cues, DNA synthesis and DNA damage (Errington, 2003). The sporulation cascade is controlled by the master regulator Spo0A (Molle et al., 2003). Phosphorylation of Spo0A is achieved via a phosphorelay system in *Bacillus* spp. that is not present in

Clostridium spp. In *C. botulinum* it is proposed that an orphan histidine kinase(CBO1120) is responsible for Spo0A phosphorylation (Wörner et al., 2006).

2.7.2 Spore structure

Bacterial spores possess a complex and co-ordinated layered structure (Fig.2.3), which contributes to their extreme resistance to a wide variety of environmental stresses including heat, radiation, desiccation and chemical (Russell, 1990; Nicholson et al., 2000). Typical layers within spores include the exosporium, spore coat, outer membrane, cortex, germ cell wall, inner membrane and in the centre of the spore the core.

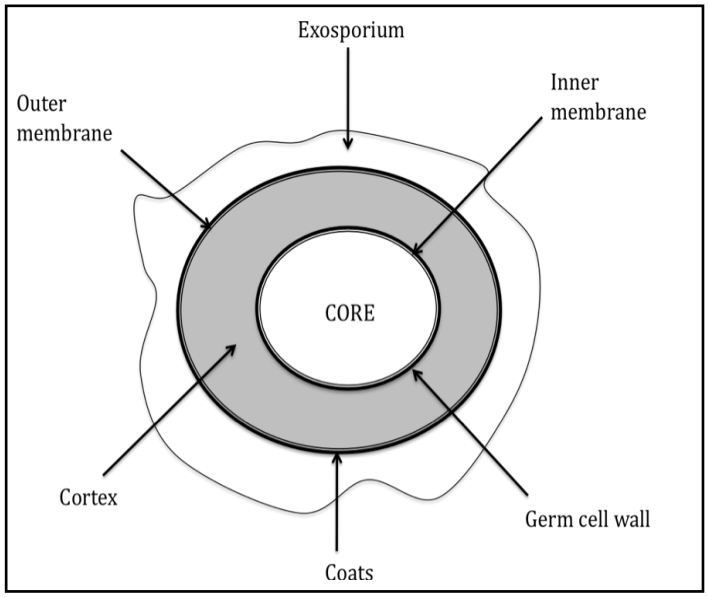

Figure 2.3 Schematic structure of spore. Note: layers are not drawn to scale**.**

The presence and morphology of both the exosporium and some surface appendages vary greatly between species. The exosporium is a sac-like structure composed predominately of water with small quantities of lipids, carbohydrates and proteins (Matz etl al., 1970). Whilst not present in all species (most notably the model sporeformer *B. subtilis*) it is found in a wide range of *Clostridium* species including *C. sporogenes* PA3679 and *C. botulinum* type E (Hodgkiss et al., 1967). The exosporium contains a variety of proteins some of which have been proposed as targets for detection methods for different spore species (Tamborrini et al., 2009). The

role of the exosporium is poorly understood however a number of potential roles include surface attachment, matrix binding of other proteins, aiding in coat assembly during sporulation and/or a protective role (Tipper and Gauthier, 1972; Ball et al., 2008).

Both within and attached externally to the exosporium, a number of bacterial species have external appendages such as ribbons, pili, feathers, brushes, tubules or swords (Rode et al., 1971; Hachisuka *et al*., 1984). The function of these appendages is relatively unexplored. It would seem quite likely that they are simply adhesive organelles however the wide diversity of types suggests they may be more specialized for survival in niche environments (Walker et al, 2007).

Beneath the exosporium is the spore coat, which consists of a multilayered proteinaceous shell. The structure and composition of the spore coat varies between species, as does the number of layers visible by electron microscopy (Driks, 1999). The spore coat of *B. subtilis* has been shown to be made up of cross-linked polypeptides and ~30 different proteins (Driks, 2007). The coat provides mechanical integrity to the spore and protects the spore from large molecules, toxic chemicals and enzymes such as lysozyme (Driks, 1999). During spore assembly the spore coat is produced within the mother cell and then directed onto the spore surface (Driks, 2007).

The spore outer membrane was initially thought to have a role during spore formation and to be highly permeable (Setlow, 2000). However, recently it has been suggested that it does provide a selective permeability barrier to small molecules including potassium ferricyanide (de Vries, 2004). The exact nature of the outer membrane still remains unknown.

The cortex is responsible for maintaining the reduced water content of the core. Unlike the germ cell wall it consists of a cross-linked peptidoglycan with specific cortex modifications. Peptidoglycan in both vegetative cells and spores of *B. subtilis* consists of alternating 1-4 linked N-acetyl glucosamine and N-acetyl muramic acid (NAM). In vegetative cells the muramic acid carries side chains consisting of either uncross-linked tripeptides or cross-linked tripeptides and tetrapetides. The spore cortex has unique structural modifications with the NAM side chains predominately consisting of the cyclic amide δ-lactam (~50%), or less commonly by a single L-

alanine (~20-30%) or uncrosslinked tetrapeptides (~20-30%), with <3% of the cortex peptidoglycan consisting of cross-linked tripeptides and tetrapeptides (Popham et al., 1996). The low degree of cross-linking is consistent with the proposed mechanism of maintaining the dehydrated state of the core by either osmotic and/or mechanical pressure exuded by the cortex (Gould and Dring, 1975). The low cross-linking varies within regions and allows the expansion of the core upon germination (Popham and Setlow, 1993).

The germ cell wall or primordial cell wall is synthesised prior to the cortical peptidoglycan and consists of what will eventually become the cell wall of the vegetative cell. Consequently it is very similar in composition to vegetative cells containing highly cross-linked and rigid peptidoglycan chains (Popham et al., 1996).

The inner membrane consists of a phospholipid and fatty acid structure not dissimilar to the inner membrane of vegetative cells; however it exists in a highly compressed state. Upon germination it expands ~2 fold without any macromolecular synthesis and assumes the role of the vegetative cell membrane. The lipids within the compressed inner membrane are immobile (Cowan et al, 2004). The inner membrane is a major permeability barrier and restricts or slows movement of hydrophilic molecules into the spore core whilst also preventing the egress of hydrophilic molecules (Setlow, 2000).

The core itself consists of the nuclear material required when the spore returns to vegetative growth including DNA, ribosomes and essential enzymes. The core exists in a highly dehydrated state with a water content of approximately 30-50% compared to the usual water content of a vegetative cell of 75-80% (Setlow, 2000). The highly dehydrated state of the core has been positively correlated with increased heat resistance in spores, thought to be due to stabilisation of proteins under dehydrated conditions. The mechanism of achieving the dehydrated state of the core during sporulation is unknown however it has been shown that the cortex is required to maintain dormancy and the dehydrated state of the core.

Within the core are also two distinctive features of the spore: i) pyridine-2,6-dicarboxylic acid or dipicolinic acid (DPA); ii) and small-acid soluble proteins (SASP). Dipicolinic acid comprises 15% of the dry weight of the core (Setlow, 2000). It is primarily found chelated with divalent ions, particularly Ca^{2+}. The role that DPA

plays within the spore is unknown however it appears to play a critical role in germination because it is immediately released upon activation. SASP's constitute up to 20% of the total spore protein and exist in α/β/γ classes. They are proteins which bind tightly to the DNA within the spore core. By binding with the DNA they tightly compact it which contributes to the resistance to UV damage (Setlow, 2007).

2.8 Thermal control of Group I *Clostridium botulinum* spores

Heating is amongst the most common methods used in food preservation. It is often the most practical and convenient method to render a product microbiologically safe and stable. The required time-temperature regimes are set based on challenge tests, legislation and/or experience (van Asselt and Zwietering, 2006). The thermal destruction of spores has traditionally been described by D-values (Stumbo, 1973). The D-value is the time taken to reduce a population by 90% at a specified temperature (T). The value assumes log-linear inactivation and may be calculated by enumeration of initial load of spores and numbers of surviving spores against time (t):

$D_T = t / \log N_o - \log N_t$.

Where N_o is the initial spore population, N_t is the surviving spores at time t.

Calculated D-values are inevitably dependent on the heating menstruum, recovery method, the strain used and preparation of the spore crop. A key assumption of D-values is that the inactivation of spores is log-linear. This, however appears to be more an exception rather than the rule, as deviations from log-linearity have frequently been observed. Whilst a number of models have been developed to try to account for the observed non-log linearity (Cerf, 1977; Peleg, 1996; Albert and Mafart, 2005; Stone et al, 2009, Corradini et al., 2010) the exact nature of the observed variability in survivor curves has never been fully elucidated and most minimum process calculations are derived by extrapolating the resultant D-values beyond the limits of the testing range.

2.8.1 Germination of spores

During dormancy the spore is in a cryptobiotic state and shows no detectable metabolic activity or ATP (Setlow, 2000). Despite being metabolically inactive spores are constantly sensing their environment via germinant receptors (Moir, 2006), and upon sensing a favourable environment the spore will break its dormancy. Spore germination is intrinsically related to the perceived heat resistance of spores. The maintenance of a dormant state provides a spore with the ability to survive extreme external stresses. This advantage however is contingent on the ability of the spore to exit this state upon sensing growth permissive environments.

The breaking of spore dormancy, may be split into three distinct events: i) activation, ii) germination and iii) outgrowth. Activation is a fully reversible process prior to germination, however it is poorly understood (Setlow, 2003). Activation is responsible for a decrease in germination lag time as well as an increase in germination rate (Curran and Evans, 1945; Powell and Hunter, 1955). In laboratories, spores are typically activated by a mild heat treatment to potentiate and thus synchronise the germination of spores (Setlow, 2003). Other activators include low pH, thiol compounds, oxidisers and calcium dipicolonate complexes (Rieman and Ordal, 1961).

In contrast to activation, germination is the irreversible change in which the activated spore becomes metabolically active (Russell, 1990). It is a degradative process where a number of biophysical changes occur including loss of heat resistance, loss of refractility, decrease in dry weight and release of dipicolinic acid (Xiao et al., 2011). Germinants, typically amino acids or purines penetrate through the outer layers of the spore coat and cortex and interact with receptors located on the inner membrane (Paidhungat and Setlow, 2001). This triggers the movement of monovalent cations from the core, followed by Ca^{2+} and DPA into the environment. This efflux allows the rehydration of the core and activation of cortex lytic enzymes which degrade the cortex to allow expansion of the core (Moir, 2003). Germination is complete after the full degradation of the spore coat and cortex. The entire process can occur within minutes under optimal conditions (Carr et al., 2010).

Upon rehydration and resumption of metabolism, molecular biosynthesis begins. Initially RNA synthesis occurs followed by protein and cell wall synthesis before

finally DNA synthesis (Russell, 1990). Eventually the spore coat is shed and a young cell emerges (Santo and Doi, 1974).

The process of spore germination in *C. sporogenes* is well described (Fujioka and Frank, 1966, Brouselle et al., 2002), however the molecular characterisation of germination related genes in spores of *Clostridium spp.* remains poorly understood compared to that of *Bacillus spp*. Variation exists in the presence, number of units and organisation of the genes associated with spore germination in *C. sporogenes* and Group I *C. botulinum* strains (Brouselle et al., 2002).

A number of the key proteins required to undergo and complete germination are localised externally to the spore core, particularly germinant receptors and cortex lytic enzymes (CLE's). As such, these proteins have been proposed as potential target sites for spore inactivation processes. There is some evidence to support the notion that these proteins are particularly sensitive to thermal treatments including increased rates of recovery of heat treated non-proteolytic *C. botulinum* in the presence of lysozyme (Scott and Bernard, 1985; Peck et al., 1992).

Sensing of nutrients within the environment by spores is mediated by germination (Ger) receptors. Whether Ger receptors work independently or may interact with other receptors to induce germination still remains unknown (Ross and Abel-Santos, 2012). Based on studies of *B. subtilis* the standard model for the structure of Ger receptors involves a protein heterocomplex consisting of three subunits (A, B and C). All three proteins are required for the formation of a functional receptor. Subunit A encodes a protein with a predicted extensive hydrophobic domain and is localised across the inner membrane of spores (Hudson et al., 2001). Subunit B is an integral membrane protein associated with germinant recognition and shares a high degree of homology to the amino-acid permease superfamily (Jack et al., 2000; Ross and Abel-Santos, 2010) and is also located in the inner membrane of spores (Alberto et al, 2005). Subunit C is a hydrophilic protein most likely anchored to the outer surface of the inner membrane (Li et al, 2010). While Ger receptors are the most well described germination initiation mechanism, it is likely that a number of alternative mechanisms of spore germination are still yet to be discovered.

2.8.2 Wet heat inactivation and resistance of spores

Whilst the heat resistance of endospores is quite well described the actual mechanism of thermal inactivation is less well understood. The proposed potential target areas are numerous including membrane damage (Flowers and Adams, 1976), key protein denaturation (Coleman et al., 2007; Coleman et al., 2010), germination system damage (as indicated above) and osmoregulatory disruption (Gould and Dring, 1975). It is likely that all of these systems are disrupted to some degree. Oxidative damage to DNA has been discounted as the primary target as moist heat treated spores have not shown any DNA damage and/or mutations, and studies with mutants harbouring deletions of DNA repair genes also demonstrated that the inability to repair DNA damage did not impact on heat resistance (Setlow, 2006).

The relationship between the release of DPA and inactivation of spores has been studied in great detail. Coleman et al. (2007) tracked the release of DPA during moist heat inactivation of *B. subtilis* and reached a number of key conclusions including: i) DPA release during heat treatments was accompanied by protein damage, ii) the release of DPA occurs either completely or not at all; iii) DPA release takes place largely if not completely after spore death and iv) spores that retain DPA can still initiate germination but not progress into outgrowth. These results were confirmed by studies on *Bacillus cereus* and *Bacillus megaterium*, (Coleman et al., 2010) suggesting it may be via denaturation of a key protein that wet heat inactivates spores.

Whilst the actual mechanism of spore inactivation is poorly understood a number of factors contributing to increased heat resistance have been identified including dehydration, mineralisation and thermal adaptation (Gerhardt and Marquis, 1989; Marquis and Shin, 1994). Spore core dehydration correlates with heat resistance, where the decreased water content is thought to stabilise proteins and DNA against heat denaturation (Atrih and Foster, 2002). Beaman and Gerhardt (1986) found that mineralisation and thermal adaption correlated with an increase in core wet weight and consequently heat resistance.

2.8.3 Sublethal injury of spores

Spores unable to outgrow after exposure to a thermal process may be characterised as either sublethally injured, or irreversibly damaged i.e. killed (Russell, 1990).

Sublethal injury of spores is a well-recognised phenomenon (Hurst, 1977; Gorman et al., 1983; Foegeding and Busta, 1983). Sublethally injured spores are unable to grow under conditions in which non-injured spores would. The term "injury" implies a temporary condition and heat damage incurred by a spore during thermal processing can be "repaired" or "recovered" during outgrowth. Heat injury has traditionally been observed by comparisons between recoveries on different culture media (Augustin and Pflug, 1967; Montville, 1984).

Unlike in vegetative cells, spore injury may manifest itself at various points including germination, outgrowth or division. Adams (1978) classified spore injury into four major classes i) requirements by survivors for non-nutrient germination stimulants, ii) modified optimum temperatures for enumeration of survivors, iii) increased sensitivity to inhibitors, and 4) altered nutritional requirements.

Inactivation of germination systems is perhaps the best studied form of injury. Spores exposed to sublethal heat injury may remain dormant unless exogenous germinants are added to bypass the primary germination system. Lysozyme in particular is often used to substitute for heat damaged cortex lytic enzymes, and has shown to provide a 1800-fold increase in recovery of non-proteolytic *C. botulinum* type E spores (Peck et al., 1992); the effect of lysozyme on proteolytic *C. botulinum* is much less indicating that the germination systems are more effectively protected by heat (Mackay, 2000). Examples of other typical clostridial non-nutrient germinants include bicarbonate (Brouselle et al., 2002), calcium dipicolinate (Riemann and Ordal, 1961) and sodium lactate in combination with L-alanine (Foegeding and Busta, 1983).

Other types of sensitivities have also been shown including by Williams and Reed (1942) who found that D-values of *C. botulinum* type A and B spores were larger when enumerated at 24ºC or 27ºC rather than at 31ºC or 37ºC. Sensitivities to inhibitory compounds in media have also been reported by Olsen and Scott (1946) who found that in heat treated *C. botulinum* spores the addition of starch in order to remove inhibitory compounds increased recovery. A proposed mechanism for the observed sensitivities is that heat may interfere with the permeability of spore membranes thus rendering the spore susceptible to inhibitory chemicals that were previously obstructed (Marquis and Shin, 1994).

2.8.4 Biovariability in spores

Traditional microbial studies have focussed on population level studies, thus limiting conclusions that can be drawn at the single cell level. As new techniques have become available it is clear that individual cells in clonal populations exhibit far more heterogeneity than previously appreciated (Avery, 2006; Veening et al., 2008).

This is especially relevant to spores (Hornstra et al., 2009; Eijlander et al., 2011), in part this is due to the number of physiological states that the spores can exist in including dormant, superdormant (Ghosh and Setlow, 2009), germinated, activated, injured or dead (Gould and Dring, 1974). A wide range of factors may influence the intrinsic genetic or phenotypic heterogeneity of the population including external stimuli such as nutrient levels, chemical germinants, physical stress, and thermal stress (Hornstra et al., 2009).

Both indirect and direct methods have been used to observe heterogeneity in spores (den Besten et al., 2007). Indirect methods are typically performed by observing optical density in single cell cultures. Smelt et al., (2008) modelled the effect of sublethal heat treatment on germination and outgrowth rates of *B. subtilis* spores. By sorting single spores into growth media and observing turbidity, they showed that a more severe treatment increased variation in lag times and showed the germination rate is dependent on prior heat treatments.

More direct methods involve imaging of individual cells such as flow cytometry or microscopy. A number of microscopic studies on the individual stages of *C. botulinum* spore outgrowth have shown that sublethal thermal treatments not only extend the lag phase but also the variability of individual stages of lag time, with an increasing mean and variation of all phases including germination, outgrowth and doubling time (Stringer et al., 2005; Webb et al., 2007; Stringer et al., 2011). Van Melis et al., (2011) also investigated the germination and outgrowth of *B. cereus* spores under sorbic acid stress. Using flow cytometry they observed a number of subpopulations under increasing sorbic acid stress which they attributed to varying stages of germination. However, despite the apparent heterogeneity they still performed a transcriptomic analysis on the entire population thus obscuring the transcript of the smaller subpopulations.

2.9 Concluding remarks

In conclusion, this review highlights the incongruence of the taxonomic separation of *C. botulinum* and *C. sporogenes* and difficulties associated with separation of these species via molecular methods. It also highlights the current explosion in understanding being made by whole genome studies on appreciating the diversity within *C. botulinum*. Currently the relationship between non-toxigenic and toxigenic strains of *C. botulinum* is poorly understood, however it is likely that a greater understanding of relationship between these strains may provide important insights into the mobility of toxigenicity within this phylogenetic unit.

In addition, this review highlighted some of the factors associated with variability in spore thermal inactivation studies of *C. botulinum* and *C. sporogenes*. As an improved mechanistic understanding of the heat inactivation of spores occurs this may lead to more efficient and novel thermal treatments that take into account secondary hurdles in food processes. However, this understanding is contingent on appreciating the impact of biovariability and state transitions in response to the thermal inactivation of spores. Further, much of the understanding of the spore life cycle is based on the model sporeformer *Bacillus subtilis* and further research on germination characteristics of *Clostridium* spp. would be of great benefit.

Subsequent chapters explore the genetics of a common non-toxigenic *C. sporogenes* (Chapter 3) and explore the genetic relationship between non-toxigenic and toxigenic strains of *C. botulinum* (Chapters 4). Furthermore, investigations into the heat resistance characteristics of spores of *C. sporogenes* with respect to biovariability (Chapter 5) and spore structure (Chapter 6) are described.

Chapter 3- Assembly, annotation and analysis of the genome of *Clostridium sporogenes* PA 3679

3.1 Introduction

The bacterial strain *Clostridium sporogenes* PA 3679 is one of the most important strains in relation to the safety and stability of thermally processed foods. The strain originally designated as Putrefactive Anaerobe (PA) 3679 was isolated from spoiled canned corn in 1927 by E.J. Cameron (as cited in Townsend et al., 1938). Since this time it has been widely employed as a non-toxigenic surrogate for Group I *Clostridium botulinum* in a broad range of food challenge studies. Its principal use is in the derivation of safe thermal processing schedules to prevent the outgrowth of Group I *C. botulinum*. The selection of *C. sporogenes* PA 3679 as a surrogate is based on three phenotypic traits: i) it produces spores that have been generally demonstrated to be more heat resistant than typical *C. botulinum* spores, ii) has growth and morphological characteristics indistinguishable from Group I *C. botulinum* strains and iii) it is non-toxigenic. The broad adoption of the strain has led to it being described by the National Committee on Microbiological Criteria for Food (NACMCF, 2010) "as an excellent surrogate" for Group I *C. botulinum* to validate thermal processes.

Despite its widespread usage *C. sporogenes* PA3679 remains poorly genetically characterised compared to *C. botulinum*, and the general identity of the strain is ambiguous (Brown et al., 2012). It is unknown whether the non-toxigenicity of the strain is dependent on the complete absence of all botulinum neurotoxin (BoNT) genes and associated non-toxic protein (ANTP's) or just a few. Previous genome-wide studies have been conducted using pulsed field gel electrophoresis (PFGE), riboprinting and comparative genomic hybridisation (CGH) against *C. botulinum* genomes. Such methods are capable of providing varying levels of discriminatory power against Group I *C. botulinum*; however, they lack the sensitivity to provide detailed information regarding the complete genetic makeup of the strain. An alternative approach is conducting whole genome shotgun sequencing (WGSS),

which may overcome these shortcomings and allow for single nucleotide resolution of the whole genome of a bacterium.

Advancement in high throughput sequencing has revolutionized the ability to conduct a WGSS on bacterial species. Current generation sequencing technologies employ parallel sequencing reactions capable of generating gigabases of sequence reads in single runs. The sequencing of genomic DNA was once considered the bottleneck for WGSS. However, it is now more often the computational requirements for dealing with larger datasets, which is considered the most intensive and difficult part of projects. A typical WGSS project for a four Mbp bacterial genome may consist of anywhere between one million to seven million reads dependent on required coverage, read length and sequencing technology. The assembly of these reads is computationally demanding as these reads may be compromised by sequencing platform dependent artefacts, systemic bias and errors. Numerous pipelines and assembly algorithms have been developed to clean errors, assemble reads and align reads (Magoc et al., 2013; Coil et al., 2014) however the area remains in a constant state of improvement.

Completion of a genome is a non-trivial task; as such the degree of completion is dependent on the experiment requirements and level of resolution required. Genomes are often described as either draft, closed/finished or polished. A draft genome is a genome that contains unresolved gaps between long stretches of contiguous sequences (contigs). These gaps typically are associated with sequences which are not resolved by the assembly algorithm such as repeated regions, tRNA's and sequence repeats. This is by far the most common approach currently employed. As of December 2013, the genomes online database (GOLD) contains 3,012 finished genomes and 15,786 permanent drafts. Closed/finished genomes are considered genomes which contain no gaps. Where once manual closing of gaps was the standard of genomes, the time, cost and technical expertise required have made this less common. Currently this requires a large insert mate-pair library, optical mapping (Latreille et al., 2007) and/or extensive primer walking. However, third generation sequencing technologies such as the PacBio RS have suggested that it is feasible to retrieve a single contig genome from a single sequencing run using multiple flow cells (English et al., 2012). A polished genome can be considered the gold standard of genomes. These genomes

have been extensively annotated with all read errors and ambiguities removed from the final genome.

This study describes a WGSS project of *C. sporogenes* PA 3679 *(*NCTC 8594) using a combination of 100 bp Illumina HiSeq and 250 bp MiSeq paired-end reads. A single contig pseudomolecule of the genome was constructed and interrogated to i) improve understanding of the identity of *C. sporogenes* PA 3679, ii) investigate potential genetic contributors to its spore heat resistance based on germination operons and iii) investigate prophage elements. Furthermore we hope that the release of the genome of *C. sporogenes* PA 3679 will benefit future studies in the selection of the strain as a surrogate for *C. botulinum*.

3.2 Materials and Methods

3.2.1 Isolate and culture conditions

An isolate of *C. sporogenes* PA 3679 was obtained from the National Collection of Type Cultures (NCTC 8594). Stock cultures were streak plated onto Reinforced Clostridial Agar (Oxoid), and incubated anaerobically at 37°C for 48 h. From this plate a single representative colony was inoculated into 10 mL tryptone-yeast extract-glucose medium consisting of 30 g L^{-1} of tryptone (Oxoid), 20 g L^{-1} of yeast extract (Oxoid), 5 g L^{-1} of glucose (Ajax) and 1 g L^{-1} of sodium thioglycollate (Ajax). This culture was incubated for a further 48 h anaerobically at 37°C before being refrigerated for 72 h prior to genomic DNA extraction.

3.2.2 DNA extraction and quality control

Genomic DNA (gDNA) was isolated with the DNeasy Blood and Tissue kit (Qiagen, Chadstone, Australia) as per the manufacturer's instructions. Resultant gDNA was assessed using a QuBit and gel electrophoresis prior to being sent to an external sequencing facility.

3.2.3 Sequencing facilities and platform

Library preparation and sequencing was performed by external facilities. Extracted gDNA was sent to the Micromon Sequencing Facility (Monash University, Clayton, Australia). Libraries were prepared using the TruSeq DNA sample preparation kit (Illumina), followed by sequencing on 1/3rd of a lane of a HiSeq platform using a 100 bp paired-end protocol.

A second sample of gDNA was extracted and sent to The Ramaciotti Centre (UNSW, Randwick, Australia). Samples were prepared for sequencing with the Nextera XT preparation kit (Illumina) and sequencing was performed on the Illumina MiSeq platform using a 250 bp paired-end protocol. The longer read length of the MiSeq were used to resolve some scaffolding errors in the preliminary HiSeq assembly.

3.2.4 Quality filtering of reads

In order to ensure stringent quality control of raw sequencing data two custom wrapper scripts were developed: *Clean_reads.sh* and *Align.sh* written in Bash shell (Appendix A). These scripts utilised a number of open-source software to perform distinct tasks. Firstly, reads deemed (by the user) to be poor quality or too short were cleaned and removed. This was followed by ordering, combining and labelling cleaned paired reads in an output suitable for downstream assembly.

As a first pass, the *Clean_reads.sh* script employed Trimmomatic v0.30 (http://www.usadellab.org/cms/index.php?page=trimmomatic) to trim reads and search for adapter contamination. The first 20 bp of all reads were removed and the last 30 bp were trimmed from each read. Clipping of any contaminating adapters along the length of the read was checked using the palindrome function against a custom file containing either the Nextera XT adapters and/or TruSeq adapters.

The resultant filtered paired end reads were then interleaved from their original two files (representing the forward and reverse reads) into a single file with any unpaired reads removed into a separate file and disregarded for downstream assembly.

Read quality was assessed pre- and post- filtering with FastQC v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

3.2.5. Assembly of Illumina paired-end reads

Two approaches were applied in the assembly of cleaned reads. Initially a *de novo* assembly of all reads was employed to derive contigs. Using these contigs a combined reference guided alignment/*de novo* hybrid assembly was conducted to generate a pseudomolecule of the *C. sporogenes* PA3679 genome.

3.2.5.1 De novo assembly

Reads from the Illumina HiSeq run were assembled into longer contigs using Velvet v1.0.0 (Zerbino and Birney, 2008) implemented on a local Galaxy server (Goecks et al., 2011). Default settings were used, with a hash length of 75 bp, with the exception of mean and standard deviation size. Optimised mean and standard deviation insert size values were derived using Picard (http://picard.sourceforge.net) after mapping against a reference genome of *C. botulinum* str. Ba4 using BWA-MEM v0.7.4 (Li, 2013).

3.2.5.2 Hybrid assembly of pseudomolecule

A pseudomolecule was created using a manual iterative approach. Assembled contigs derived from the *de novo* assembly were ordered against *C. botulinum* str. Ba4 using progressiveMauve (Darling et al., 2010). These contigs were then concatenated and 250 bp reads were mapped against the concatenated sequence using BWA-MEM v0.7.4. The alignment was manually assessed by visualisation in Geneious (v6.1.4; http://www.geneious.com/) to identify any ambiguities between the reads and the concatenated sequences. Contigs were then extended based on consensus of reads and orientation of paired ends. This process was repeated until ambiguities could not be unequivocally resolved. The final pseudomolecule was concatenated with a 10 bp 'N' gap placed between contigs.

3.2.6 Annotation and COG calling

Annotation of the genome was performed using Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/). Functional annotation of coding sequences was conducted using the Integrated Microbial Genomes server (Markowitz et al., 2012).

3.2.7 Visualisation of pseudomolecule

The genome ring of *C. sporogenes* was visualised using the Gview server using default settings (Available from https://server.gview.ca/) (Petkau et al., 2010). Pairwise comparison of sequences were visualised using WebACT (Abbott et al., 2005).

3.2.8 Phylogenetic analysis of *C. sporogenes* PA 3679

A phylogenetic comparison of the 16S rRNA of *C. sporogenes* PA3679 was conducted against a combination of neighbour joining trees of the aligned sequences generated with Geneious v6.1.4 using the Tamura and Nei (1993) model parameter model with gamma correction and 1,000 bootstrap replicates for all sequences.

3.2.9 Phage identification

Intact and partial putative prophage were called using PHAST (Zhou et al., 2011).

3.2.10 NCBI Accession Number

The initial 107 contigs generated from this Whole Genome Shotgun project have been deposited at DDBJ/EMBL/GenBank under the accession no.AGAH00000000.

3.3 Results

3.3.1 Pre-assembly processing and analysis of Illumina paired-end data

Prior to assembling reads, it was necessary to ensure that reads were appropriately cleaned and characterised to maximise the assembly efficiency. In total 8,178,972 HiSeq reads of length 100 bp and 429,959 paired-end MiSeq reads of length 250 bp were used to construct a draft assembly of the *C. sporogenes* PA 3679 genome. Based on an estimated genome size of 4.0 Mbp this equated to a predicted coverage over forty times.

Quality assessment of reads was made using FastQC. Due to the high number of reads stringent quality control was employed. The processing of reads led to >25% of reads being removed, but resulted in improved uniformity and quality. Example graphs generated by FastQC as shown in Figures 3.1 and 3.2 highlight the difference in quality scores and nucleotide content per bp between pre-processed and quality filtered reads. Pre-processed reads exhibited typical characteristics of reads generated using Illumina MiSeq and HiSeq systems (Figures 3.1 and 3.2). The most notable characteristic is the higher quality scores at the start of the reads (5' end) followed by a decrease in quality towards the end (3' end). Another typical characteristic is the significant variability exhibited by the first 15-20 bp, which is commonly associated with adapter contamination of reads. Both of these characteristics significantly decrease the efficiency of downstream *de novo* assemblies. The application of the *Clean_reads.sh* script removed both of these potential errors from the reads and significantly improved the assembly.

The insert size of reads (the distance between corresponding pairs) was assessed using Picard (Figure 3.3). The calculated median insert length for the HiSeq and MiSeq reads were 231±22 and 278±61 respectively. Using a TruSeq library preparation kit, for HiSeq, which uses a size selection step resulted in more uniform distribution compared to Nextera XT, for MiSeq, with enzymatic based distribution and no overlapping paired ends.

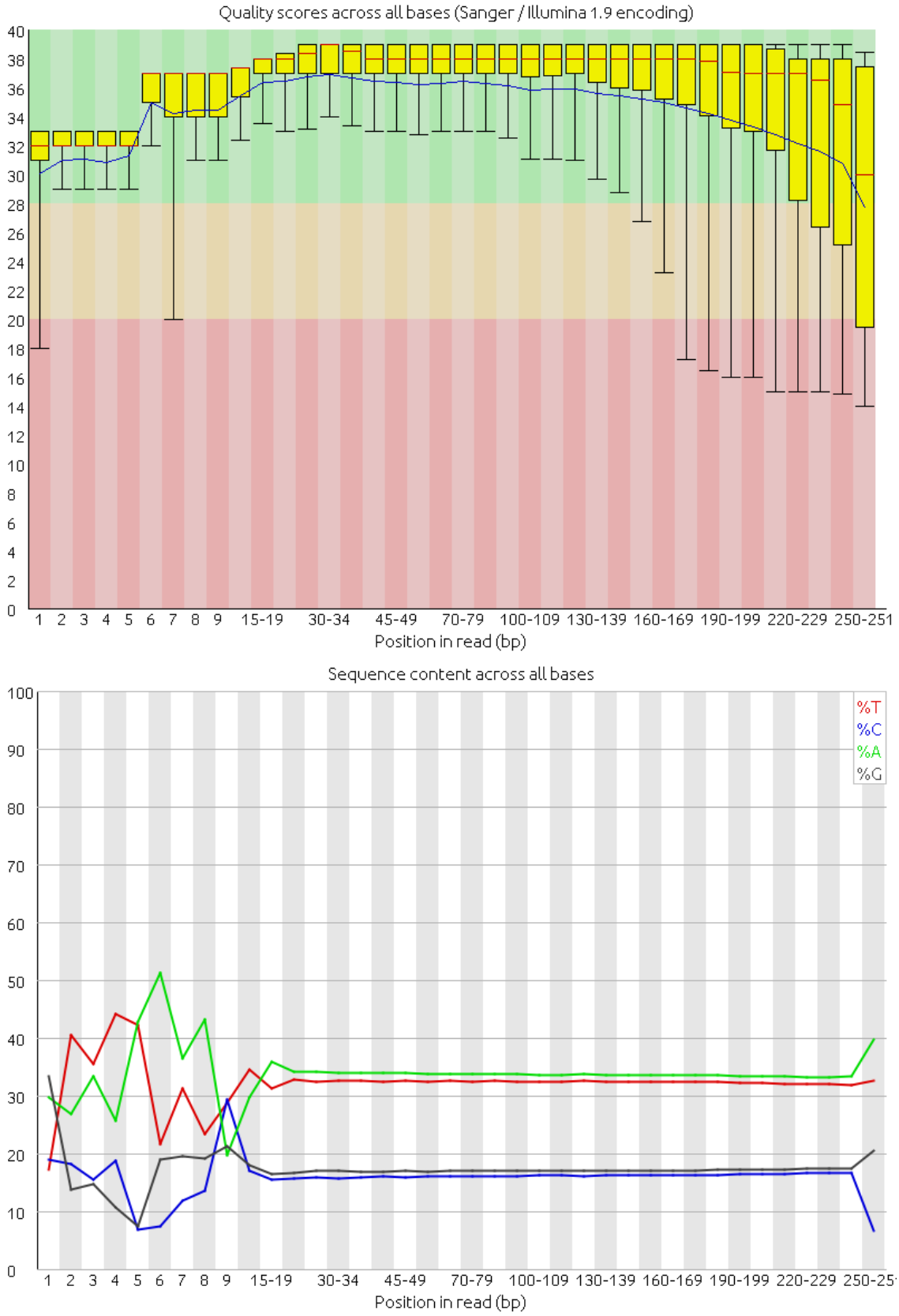

Figure 3.1 Read quality assessment produced by FastQC for 250 bp MiSeq reads of *C. sporogenes* PA 3679 prior to processing using *Clean_reads.sh* script. The top graph represents the median quality scores at each base position along reads. The bottom graph represents the diversity of individual nucleotides at each base position.

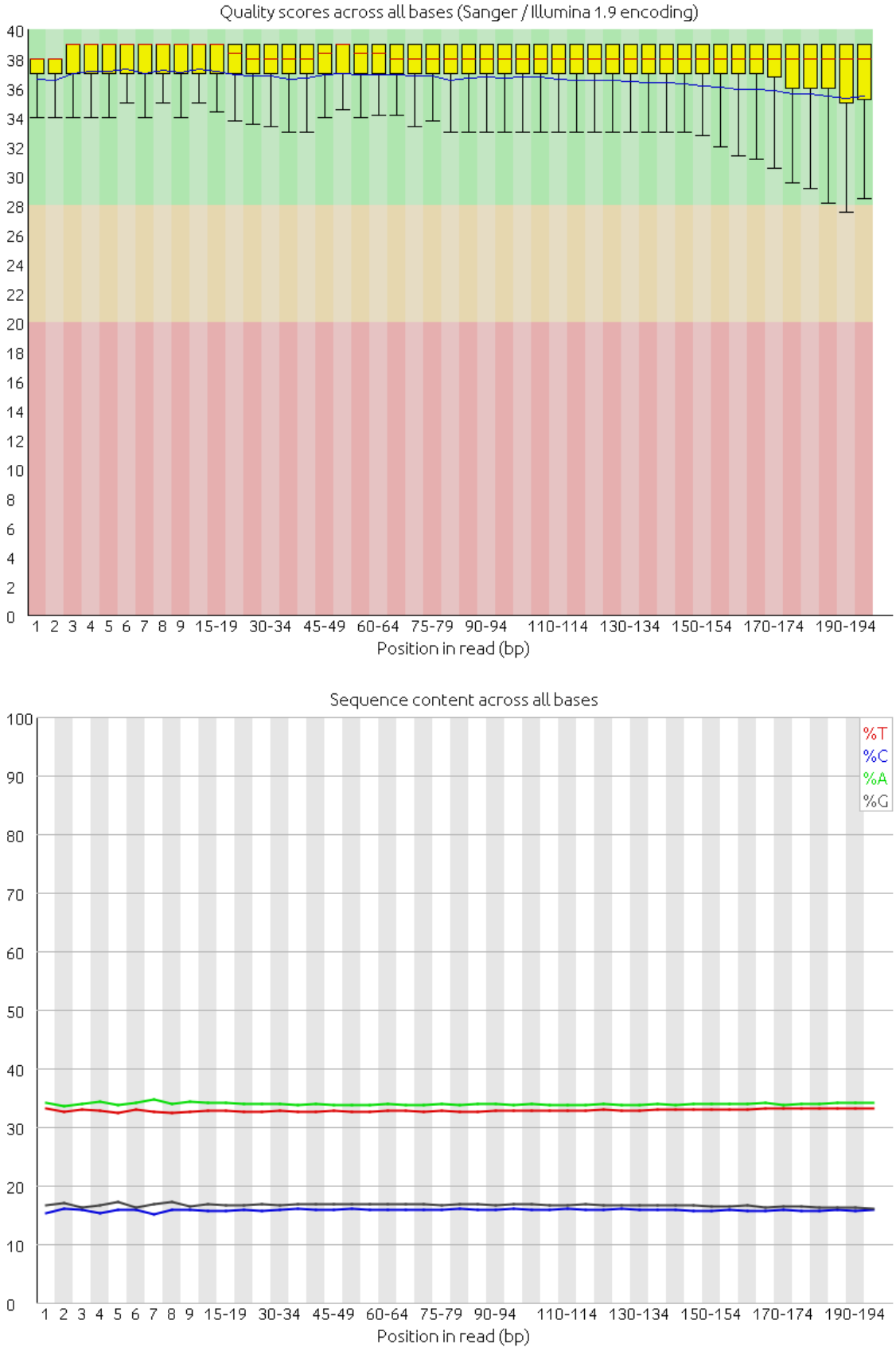

Figure 3.2 Read quality assessment produced by FastQC for 250 bp MiSeq of *C. sporogenes* PA 3679 post processing using *Clean_reads.sh* script. The top graph represents the median quality scores at each base position along reads. The bottom graph represents the diversity of individual nucleotides at each base position.

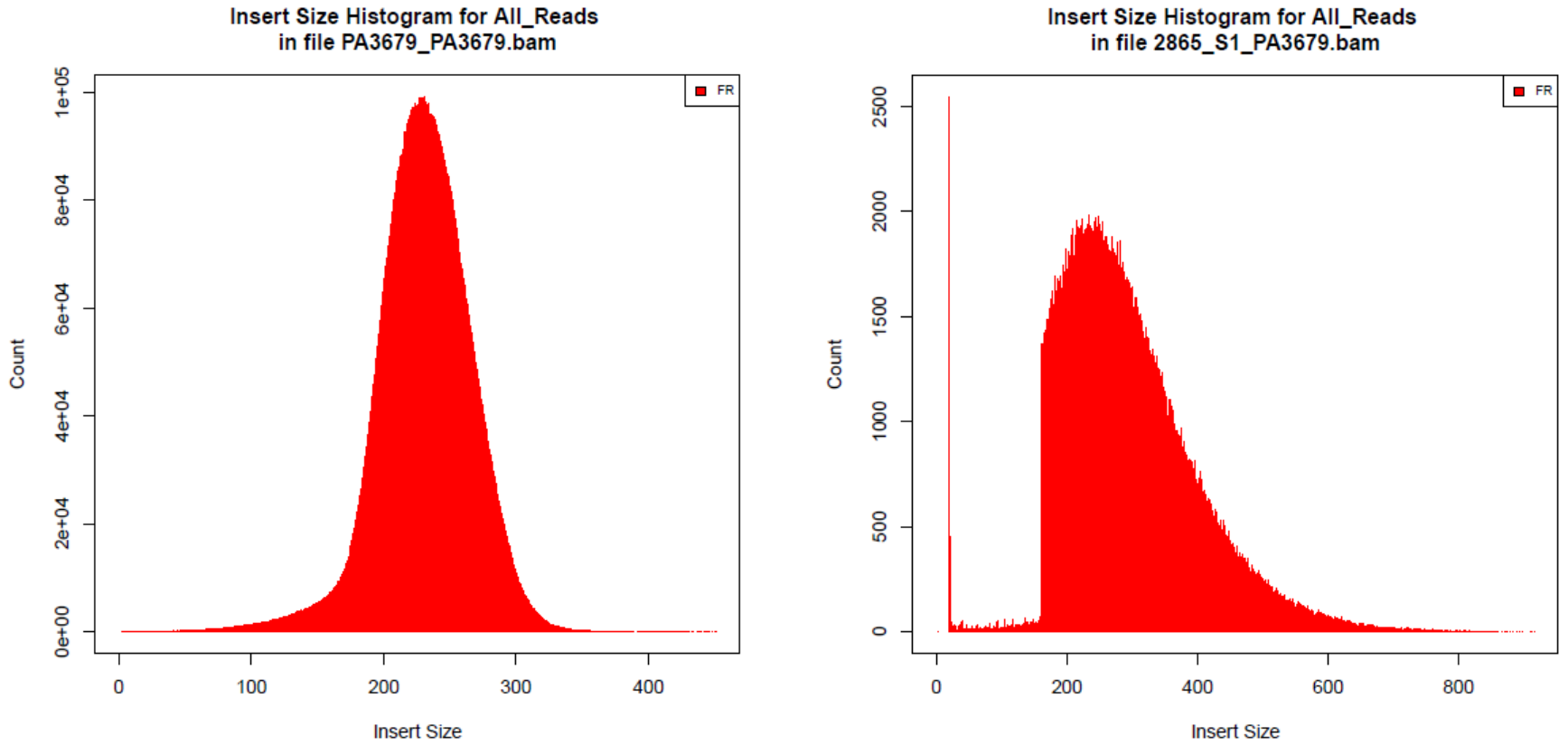

Figure 3.3 Differences in insert size distributions between a 1/3rd a lane of 2 x 100 bp HiSeq with library preparation using the TruSeq library preparation kit (left) in comparison to a 2 x 250 bp MiSeq run prepared with a Nextera XT preparation kit (right). Note: An insert size of zero represents overlapping reads.

3.3.2 Draft assembly of *C. sporogenes* PA 3679

The preliminary assembly output produced 107 contigs representing the genome of *C. sporogenes* PA3679 (Table 3.1). A combination of secondary assembly and read mapping decreased the number of contigs to 41. The maximum contig length was 788 kbp and greater than 50% of the total genome (N50) was represented by contigs longer than 236 kbp in size.

Table 3.1 *De novo* assembly statistics of preliminary assembly using Velvet v1.0.0 and final assembly after manual iterative expansion of contigs

	Preliminary assembly statistics	**Contigs after manual iteration**
Number of contigs	107	41
Maximum contig length	425,355	788,177
N50	172,248	236,178
N90	36,883	50,571

The resulting contigs were ordered by synteny to the genome of *C. botulinum* str. Ba4 (the largest genome within this phylogenetic unit) (Figure 3.4). All contigs not sharing synteny to this genome were reordered by size at the end of the assembly. The vast majority of contigs (>90%) showed some synteny to the *C. botulinum* genomes. Those contigs that did not show synteny were all less than 200 bp in length and contained either RNA or phage associated elements. The contigs were concatenated with a 10 bp 'N' gap inserted to produce a pseudomolecule representing the final genome.

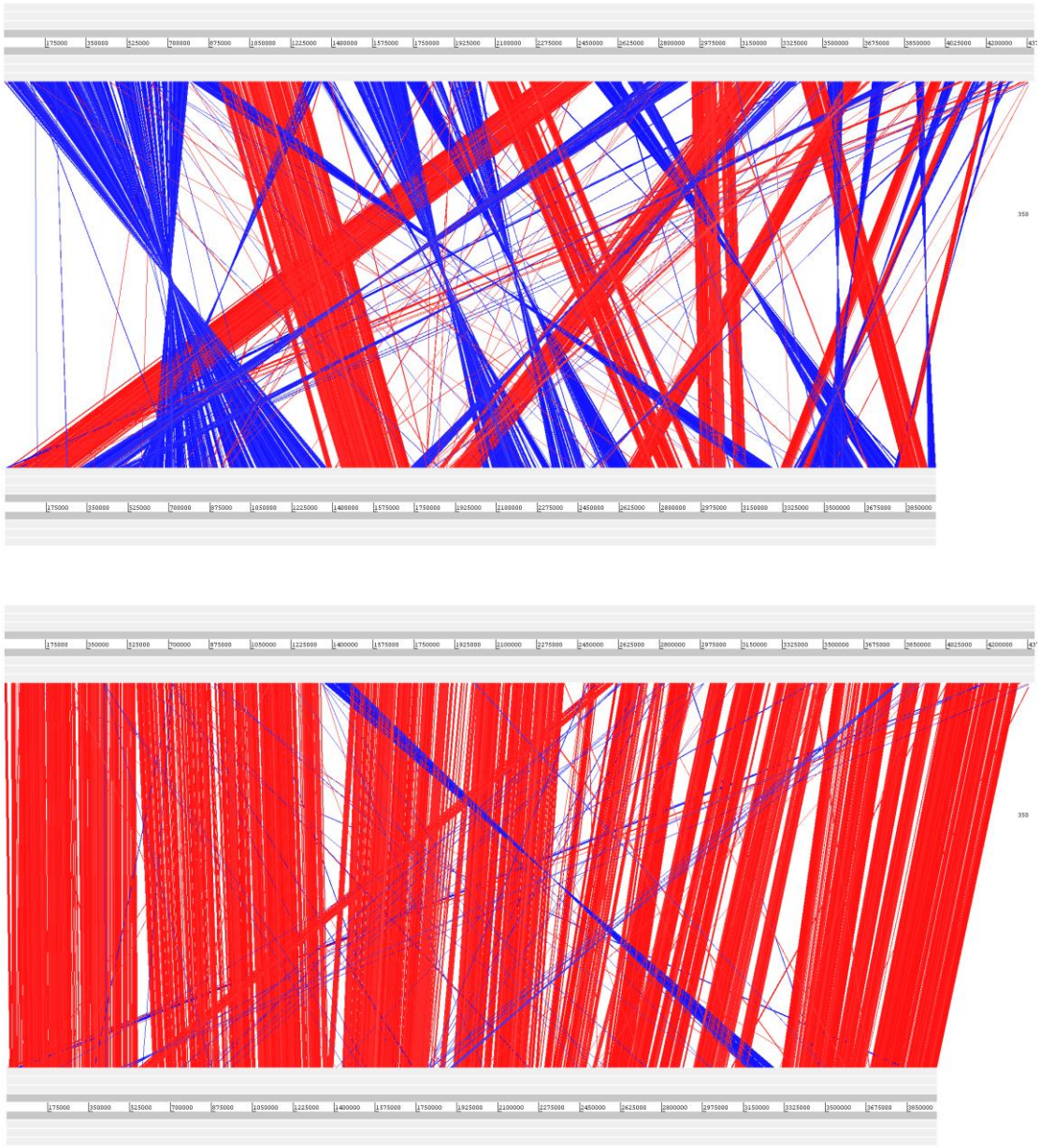

Figure 3.4 Artemis Comparison Tool nucleotide comparisons of unordered (a) and ordered (b) contigs generated from the *de novo* assembly of *C. sporogenes* PA3679 (NCTC 8594) against a reference *C. botulinum* str. Ba4 genome using progressiveMauve. The top line of each comparison represents contigs, whilst the bottom line represents the reference genome. Regions with similarity are highlighted by connected red or blue lines between the genomes; red lines indicate homologous blocks of sequence and blue lines indicate inversions. Gaps indicate unique DNA.

3.3.3 Identification of high coverage nodes in assembly

A key assembly parameter is coverage which represents the number of reads assembled at a single nucleotide position. The average coverage of 88 of the preliminary 107 contigs is shown in Figure 3.5. Three contigs (2, 4 and 19) had a

coverage level that was 60-fold greater than other contigs. These contigs were used for a second assembly with the assembly parameters altered to an estimated coverage of 1000 with the removal of nodes with low coverage (less than 500). Based on this assembly a 38.6kbp contig was extracted representing a complete bacteriophage genome (see section 3.3.7). Despite representing less than 1% of the total genome the contig was generated by ~28.4% of total reads used in this assembly, representing a large number of copies.

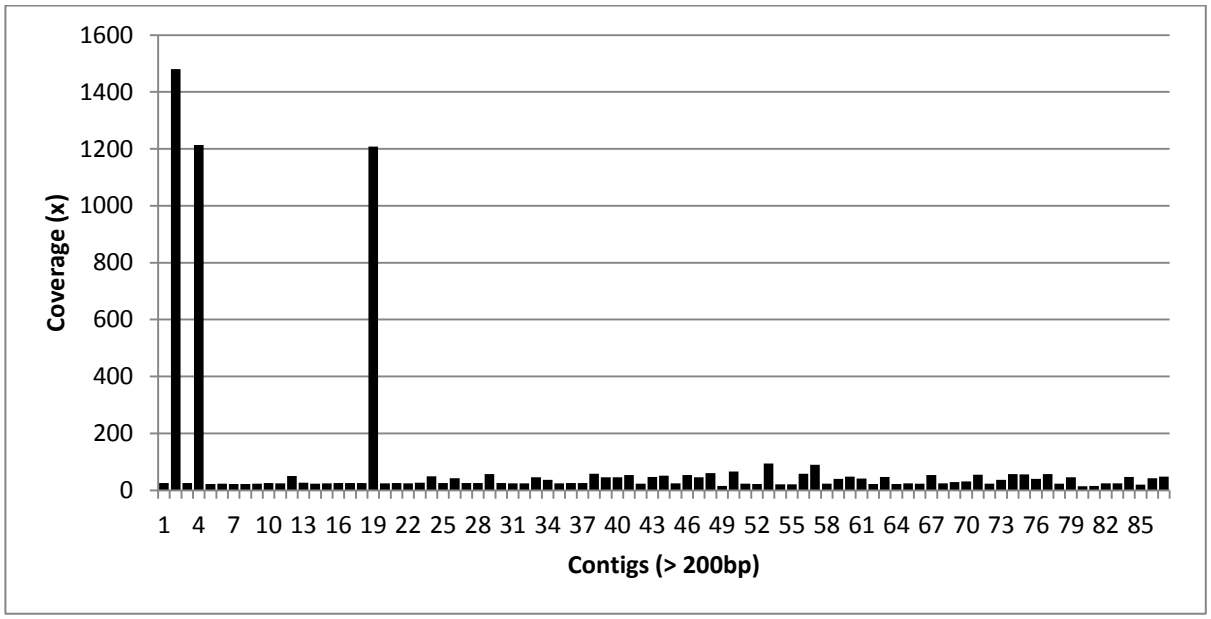

Figure 3.5 Coverage of contigs derived during preliminary *de novo* assembly of the Illumina reads of *Clostridium sporogenes* PA 3679.

3.3.4 General characteristics of the genome of *C. sporogenes* PA 3679

The total size of the draft genome of *C. sporogenes* PA 3679 (NCTC 8594) was 4,180,273 bp with a GC content of 27.79%. The general features of the genome are summarized in Table 3.2. Accurate quantification of the number of RNA genes was unable to be performed due to ambiguities in assembly generated by the copy number of these genes and sequence similarity between RNA genes. However 43 tRNA's and 31 other RNA's were identified.

Visualisation of the genome pseudomolecule (Figure 3.6) revealed localised GC content spikes across the genome. The GC skew of the genome allowed the origin of replication and terminus of replication to be clearly defined at the start of the genome

and at ~2 Mbp, respectively by a change from green to purple. Consistent with GC skew was the directional bias in the strand of genes with the large majority of those genes in the front proportion of the genome belonging to the forward strand (blue) and the majority of other genes belonging to the reverse strand (orange).

Table 3.1 Genomic features of the *C. sporogenes* PA 3679 genome and comparison with the draft genome of *C. sporogenes ATCC 15579* and representative Group I, Group II and Group III *C. botulinum* genomes. Statistics were derived using IMG.

	C. sporogenes **PA3679**	***C. botulinum*** **ATCC 3502**	***C. sporogenes*** **ATCC 15579**	***C. botulinum*** **Eklund 17B**	***C. botulinum*** **BKT015925**
Source:	Canned corn, USA	Canned peas, USA	Human faeces	Marine sediment, USA	Poultry outbreak, Sweden
Project status	Draft	Complete	Draft	Complete	Complete
Total genome size (bp)[a]	4,180,273	3,903,260	4,102,125	3,847,969	3,207,592
G+C content (%)	27.79	28.24	28.02	27.48	28.23
Total number of genes	4111	3825	3869	3639	3109
Coding area (%)	98.15	82.55	97.36	96.9	96.3
RNA operons	76[b]	89	102	112	115
Predicted prophage in chromosome[c]	10	3	2	6	2

[a] Genome size includes chromosome and plasmid

[b] RNA operons in *C. sporogenes* PA 3679 are a prediction based on coverage and were not unambiguously identified

[c] Predicted prophage (including intact and incomplete) were identified using PHAST.

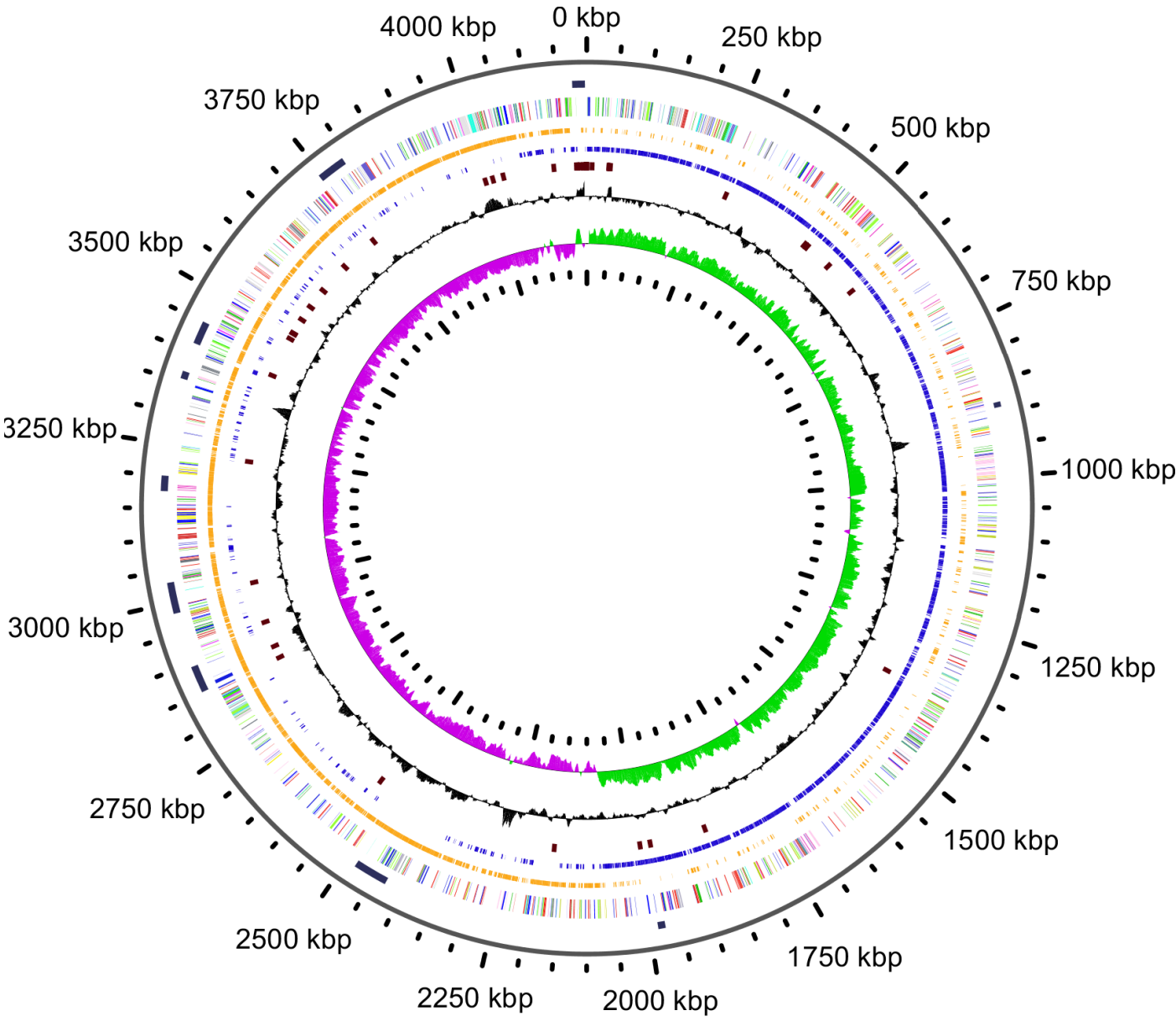

Figure 3.6 Genome ring of the *C. sporogenes* PA 3679 (NCTC 8594) pseudomolecule. The rings, listed from the inner to outer tracks are as follows: The inner ring represents the GC skew with green representing a positive skew and purple representing a negative skew. The black track represents GC content. The burgundy markers represent the location of unresolved gaps between contigs represented by a 10 bp 'N's. The blue ring represents positive strand genes. The orange ring represents negative strand genes. The coloured bars represent the COG grouping of genes assigned to COG categories, the colour of which corresponds to Table 3.2. The dark blue ring is prophage regions identified using PHAST. The outer ring is a genome backbone.

Table 3.2 COG functional classifications of genes located in the genome of *C. sporogenes* PA 3679 (NCTC 8594)

COG Categories	Gene Count
Amino acid transport and metabolism (E)	288
Carbohydrate transport and metabolism (G)	143
Cell cycle control, cell division, chromosome partitioning (D)	40
Cell motility (N)	90
Cell wall/membrane/envelope biogenesis (M)	182
Chromatin structure and dynamics (B)	1
Coenzyme transport and metabolism (H)	135
Defense mechanisms (V)	114
Energy production and conversion (C)	182
Function unknown (S)	305
General function prediction only (R)	387
Inorganic ion transport and metabolism (P)	175
Intracellular trafficking, secretion, and vesicular transport (U)	48
Lipid transport and metabolism (I)	57
Nucleotide transport and metabolism (F)	84
Posttranslational modification, protein turnover, chaperones (O)	92
Replication, recombination and repair (L)	163
RNA processing and modification (A)	1
Secondary metabolites biosynthesis, transport and catabolism (Q)	34
Signal transduction mechanisms (T)	229
Transcription (K)	333
Translation, ribosomal structure and biogenesis (J)	180
Not in COGs	1175

3.3.5 Phylogeny of *C. sporogenes* PA 3679

The broad phylogenetic positioning of *C. sporogenes* PA 3679 was identified by extracting the complete 16S rRNA gene and preparing a phylogenetic tree using all completed genomes from the genus *Clostridium* available in the NCBI genome database (as of December, 2013). The resultant tree (Figure 3.7) identified the positioning of *C. sporogenes* PA 3679 in the same clade as Group I *C. botulinum*. This clade was significantly closer to a number of other species compared to clades containing Group II and Group III *C. botulinum* strains.

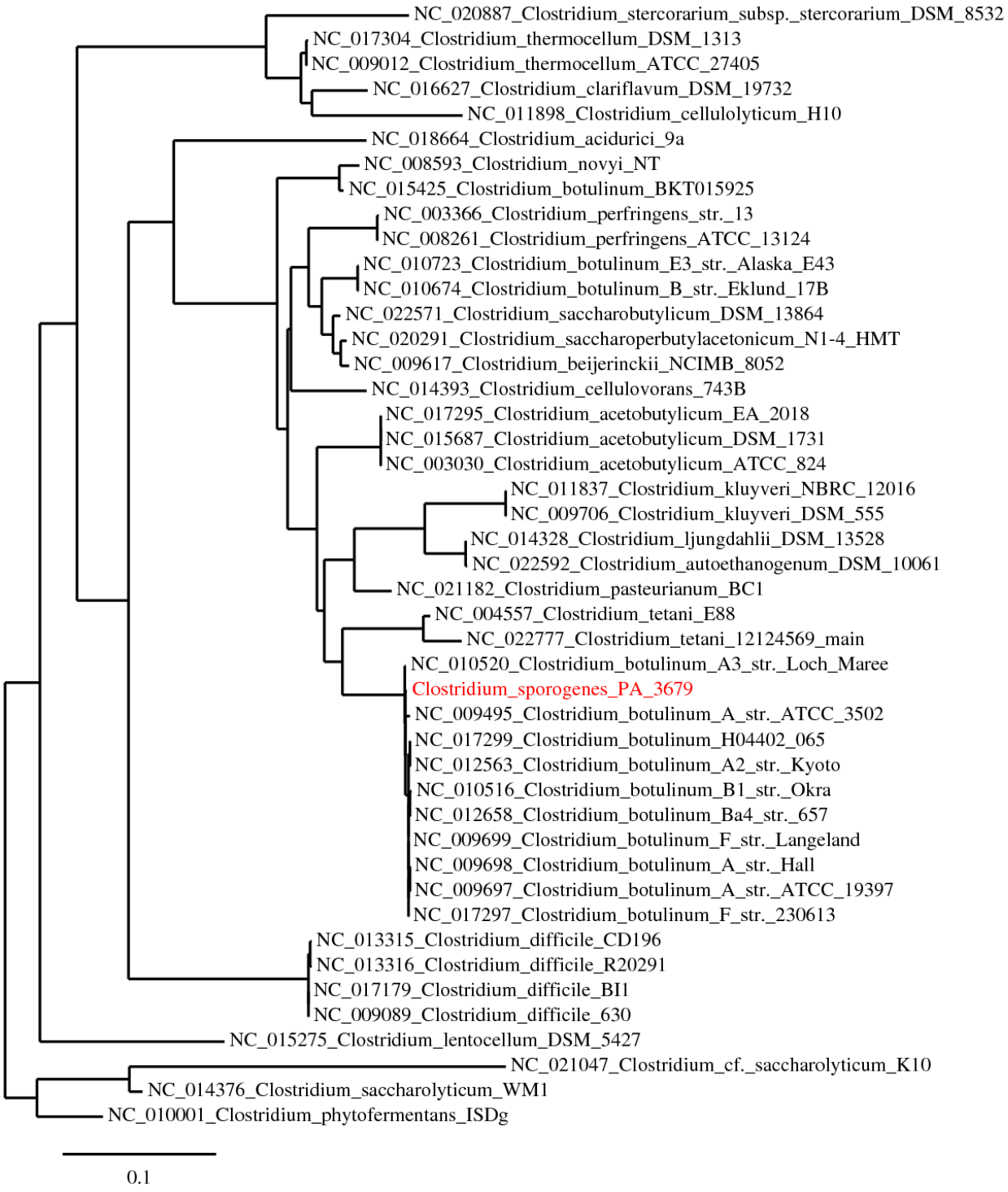

Figure 3.7 Phylogenetic tree of the complete 16S rRNA of *C. sporogenes* PA 3679 referenced against completed bacterial clostridia sequences located in the NCBI genome database (Accessed: December, 2013).

3.3.6 Functional and phylogenetic prediction of genes of *C. sporogenes* PA 3679

A total of 4111 genes were called (Appendix B), with 4035 of these representing protein coding genes. Functional characterisation of genes was conducted by classification of genes into characterisation of genes (COG) groups, which represent conserved protein domains. In total 2936 genes were classified in COG groups as described in Table 3.3. Of these the largest numbers of genes were associated with general function prediction (387), transcription (333), function unknown (305) and amino acid metabolism (288).

The phylogenetic relationships of protein coding genes present in the *C. sporogenes* PA 3679 genome were inspected for potential horizontally transferred components and novel features. The BLASTN similarity of protein coding genes was conducted at 30%, 60% and 90% levels using the IMG database containing, as of December 2013, 12509 bacterial genomes.

At the 90% similarity level over 20% of the genes could not be assigned whilst the remaining genes were assigned to the class *Clostridia*, except one gene, encoding a ribosomal protein L36 (locus tag IYC_15348), which was assigned to the class *Bacilli*. When the similarity index cut-off was lowered to 60% more than 90% of genes were assigned to the class *Clostridia* increasing to over 95% as the similarity index cut-off was decreased to 30%. Only 89 protein coding genes remained unassigned at the 30% level. These genes encoded hypothetical proteins and were predominately found within or near phage elements.

Table 3.3 Phylogenetic distribution of protein coding genes of *C. sporogenes* PA 3679 (n=4035) to different classes based on BLAST similarity at 30%, 60% and 90%. Similarity search was conducted using the IMG database.

Phylum/Class	No. of Genomes	No. Of Hits 30%	% Hits 30%	No. Of Hits 60%	% Hits 60%	No. Of Hits 90%	% Hits 90%
Euryarchaeota	258	1	0.02%				
Actinobacteria	1321	1	0.02%				
Bacteroidetes	516	6	0.15%	2	0.05%		
Chloroflexi	32	1	0.02%				
Bacilli	2548	63	1.56%	21	0.52%	1	0.02%
Clostridia	722	3847	95.34%	3667	90.88%	3091	76.60%
Erysipelotrichi	25	5	0.12%	1	0.02%		
Negativicutes	76	6	0.15%	4	0.10%		
Fusobacteria	50	3	0.07%	1	0.02%		
Alphaproteobacteria	1038	1	0.02%				
Deltaproteobacteria	145	1	0.02%				
Gammaproteobacteria	4120	5	0.12%				
Spirochaetes	412	1	0.02%				
Unclassified bacteria	279	3	0.07%	3	0.07%		
Neocallimastigomycota	2	1	0.02%				
dsDNA viruses, no RNA stage	965	1	0.02%	1	0.02%		
Unassigned	-	89	2.21%	335	8.30%	943	23.37%

3.3.7 Prophage regions identified in *C. sporogenes* PA 3679

Four intact and six incomplete phage regions were identified using PHAST in the genome of *C. sporogenes* PA3679 (Table 3.4). The size of regions ranged from ~8 kbp to ~55 kbp. The regions were identified as Clostri_PA3679_φ01 - Clostri_PA3679_φ10. Overall, the location and placement of the majority of phage were in the second half of the genome.

Table 3.4 Prophage regions identified using PHAST in *C. sporogenes* PA 3679.

REGION	LENGTH	COMPLETENESS	REGION_POSITION	POSSIBLE PHAGE	GC_%
Clostri_PA3679_φ01	8.3Kb	incomplete	881625-890021	PHAGE_Ostreo_OlV1_NC_014766, ...	26.8%
Clostri_PA3679_φ02	11.7Kb	incomplete	1975257-1986974	PHAGE_Clostr_phi_CD119_NC_007917, ...	29.4%
Clostri_PA3679_φ03	54.8Kb	intact	2429524-2484359	PHAGE_Geobac_virus_E2_NC_009552, ...	30.4%
Clostri_PA3679_φ04	42.7Kb	intact	2851503-2894268	PHAGE_Psychr_pOW20_A_NC_020841, ...	30.9%
Clostri_PA3679_φ05	49.5Kb	intact	2982116-3031703	PHAGE_Clostr_phi8074_B1_NC_019924, ...	28%
Clostri_PA3679_φ06	23.3Kb	incomplete	3176351-3199699	PHAGE_Cafete_BV_PW1_NC_014637, ...	26.1%
Clostri_PA3679_φ07	12Kb	incomplete	3355361-3367444	PHAGE_Feldma_species_virus_NC_011183, ...	31%
Clostri_PA3679_φ08	36.5Kb	incomplete	3414201-3450782	PHAGE_Clostr_phiC2_NC_009231, ...	28.8%
Clostri_PA3679_φ09	45.6Kb	intact	3744606-3790212	PHAGE_Clostr_phi_CD119_NC_007917, ...	29.3%
Clostri_PA3679_φ10	20.2Kb	incomplete	4175718-4196005	PHAGE_Clostr_phi8074_B1_NC_019924, ...	31.9%

As described in section 3.3.3, one phage region (Clostri_PA3679_φ8) was identified via higher average contig coverage. Despite PHAST identifying this phage as incomplete, the high coverage is indicative of multiple copies and induction of this phage. A BLASTN search of Clostri_PA3679_φ8 revealed significant similarities between this phage and a region present in *C. botulinum* str. Kyoto. Comparison between the prophage revealed similarities in the structural components of the phage genomes (Figure 3.8), in particular those genes encoding head-tail and tail proteins. There was however, little similarity in the rest of the phage genomes excluding a small region responsible for lysogeny control.

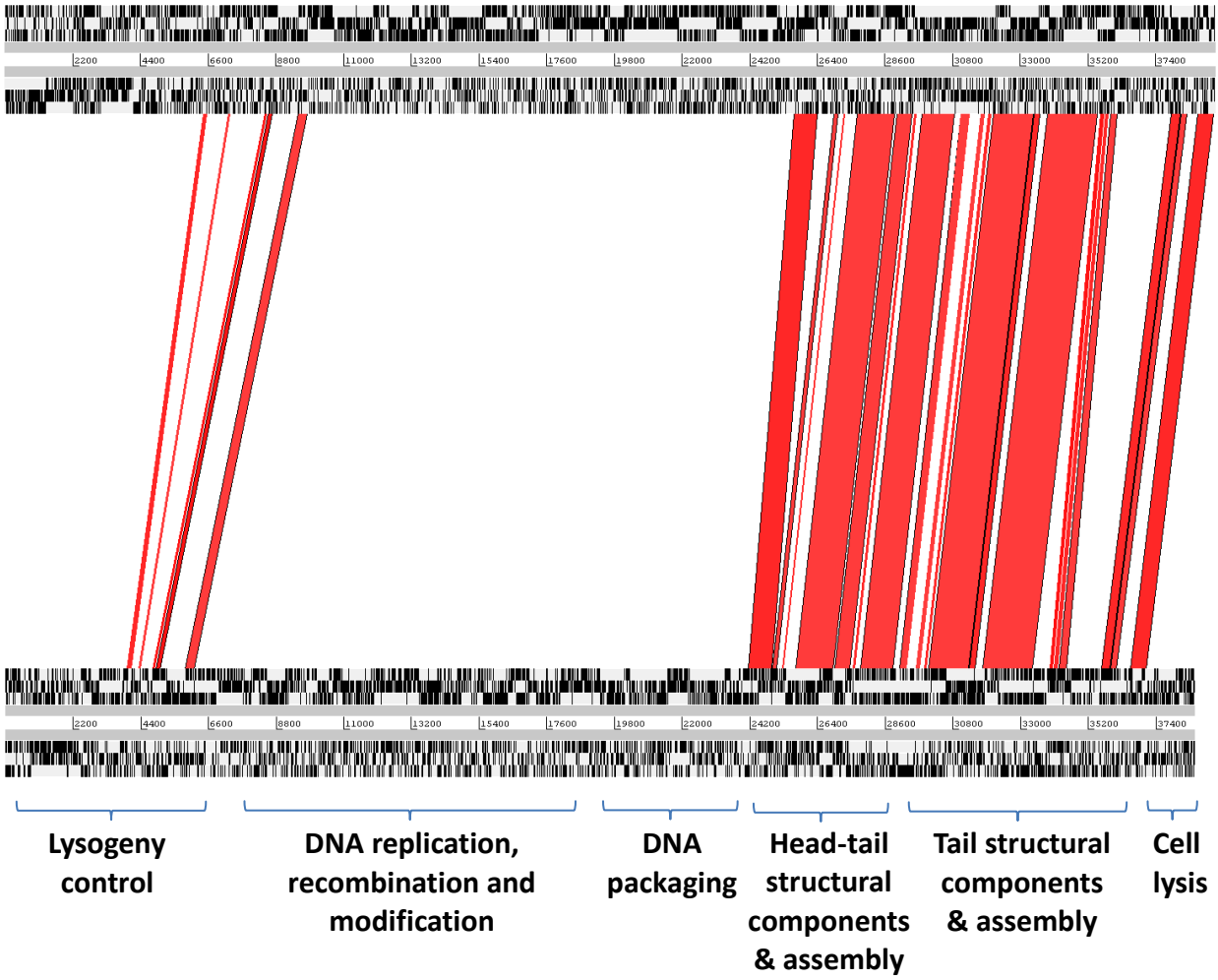

Figure 3.8 Artemis Comparison Tool nucleotide comparisons of *C. botulinum* str. Kyoto prophage (top) and Clostri_PA3679_φ8 present in *C. sporogenes* PA 3679 (bottom).

3.3.8 Germination operons in *C. sporogenes* PA 3679

As the spore characteristics are an important phenetic trait of *C. sporogenes* PA 3679, the genome was inspected for germination related operons and coding sequences. The structure of Ger receptors consists of three subunits (A, B and C), and all three are required for a functional receptor. The genome contained five polycistronic germination operons arranged as BABCB, ABC, ABC, ABC and BAC (Figure 3.9).

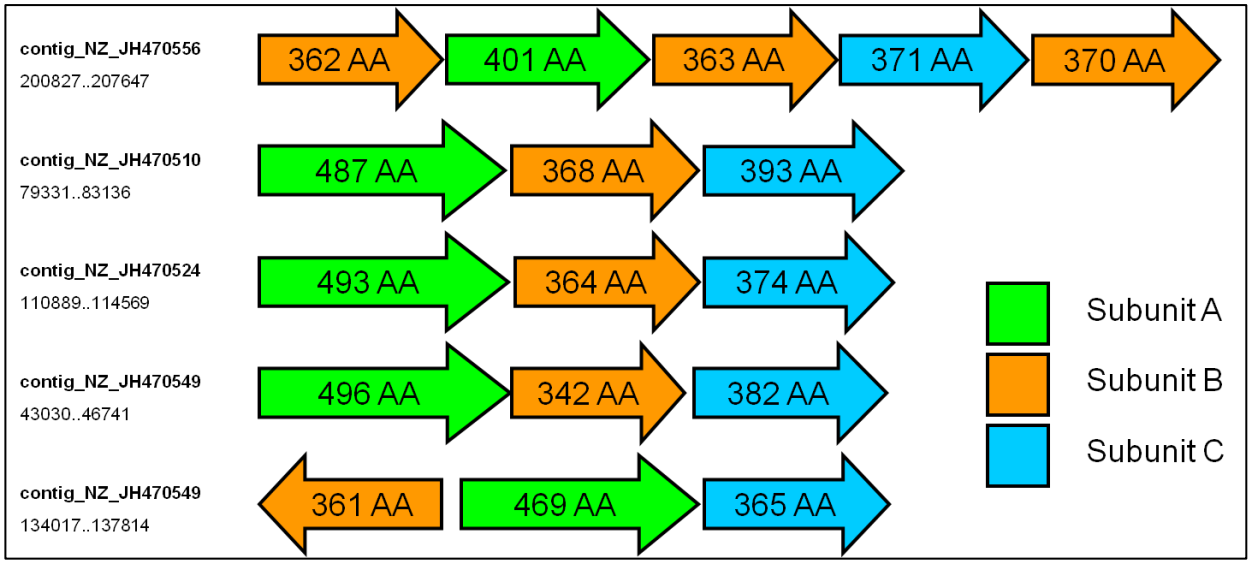

Figure 3.9 Diagrammatic representation of polycistronic germination operons identified in the draft assembly of *C. sporogenes* PA3679 (NCTC 8594). Each subunit was labelled as subunit A, B or C. Direction is based on synteny to completed genomes of *C. botulinum*.

Other genes associated with environmental sensing of spores observed in the draft assembly include an orphan *GerB* gene located downstream of the pentacistronic BABCB operon. Genes encoding homologs of PrkC and a Ser/Thr Kinase (IYC_10484) were also identified. Homologs of genes encoding cortex lytic enzymes required for cortex hydrolysis in *Bacillus spp*. spores including SleB (IYC_08060), cwlJ (IYC_20065) and YpeB (IYC_07635) were also identified in *C. sporogenes* PA 3679 indicative of similar germination mechanism as in *Bacillus spp*.

3.4 Discussion

Whilst a number of genomes of Group I *C. botulinum* have been sequenced (Sebaihia et al., 2007; Dover et al., 2014) previously only one *C. sporogenes* was available on the NCBI database. This work constructed a draft genome of *C. sporogenes* PA 3679 which was used to derive a number of conclusions about the strain. In order to produce this genome a pipeline that provided a single, annotated pseudomolecule from raw Illumina data without the use of optical mapping or long mate pairs was developed. From the resultant pseudomolecule significant conclusions regarding the nature of the genome including the presence of a large number of prophage within the genome, evidence for a hitherto described temperate phage and the presence of a number of germination genes were able to be drawn.

3.4.1 Construction of pseudomolecule from raw Illumina reads

The contigs produced solely by employing an assembly algorithm were significantly improved by manual growing of contigs. The number of contigs was halved from 107 to 41 using manual approaches. A possible weakness of the method employed is inversions that could be present in the genome of *C. sporogenes*; however the GC skew clearly identified leading and lagging strands in both halves of the pseudomolecule representing *C. sporogenes.* This is further confirmation that the assembly accurately represents the genome. Increasingly a number of bioinformatics tools such as A5 MiSeq pipeline (Coil et al., 2014) and PAGIT (Swain et al., 2012) have been developed which are likely to further improve assemblies derived using traditional assembly algorithms such as Velvet, which was used here.

3.4.2 The genome of *C. sporogenes* reveals synteny with Group I *C. botulinum* genomes

A number of notable features were observed in the genome of *C. sporogenes* PA 3679. It was larger than most *C. botulinum* genomes and showed significant synteny with *C. botulinum* str. Ba4 (Figure 3.4). The general characteristics of the genome related well to Group I *C. botulinum*. The phylogenetic analysis of the 16S rRNA gene confirmed that *C. sporogenes* belonged to the same clade as *C. botulinum*. The genome exhibited the majority of the same genetic characteristics as these genomes

consistent with a similar saprophytic life cycle and a niche adaption for survival in soil environments through spore formation.

3.4.3 *C. sporogenes* has distinct germination operons

Germination related genes of *C. sporogenes* PA 3679 were explored as a number of these proteins are required to undergo and complete germination and typically localise external to the spore core. As such, these proteins have been proposed as potential target sites for spore inactivation processes. The genome contained five polycistronic germination operons, a higher number than any of the currently sequenced draft or complete genomes of group I *C. botulinum* or *C. sporogenes*. The specificity of germinants previously associated with *C. botulinum* and *C. sporogenes* include simple amino acids, sugars and purine derivatives.

The structure of germination operons has been shown to be relatively well conserved. Studies of *Bacillus subtlis* the standard model for the structure of Ger receptors, identified a protein heterocomplex consisting of three subunits (A, B and C). All three proteins are required for a functional receptor. A genome mining study by Xiao et al., (2011) has split the Group I *C. botulinum* into two groups, i) those containing two complete tricistronic *ger* operons and a pentacistronic *ger* operon (a BABCB operon and two ABC operons) or ii) those in addition containing an extra BAC divergon with an organisational similarity to non-proteolytic *C. botulinum*. The presence of the BAC operon has been proposed to potentially increase the range of germinants (Peck, 2009); however the exact specificity of the receptor encoded by this operon is unknown.

This assembly of *C. sporogenes* PA 3679 contained a third ABC operon downstream of the BABCB operon. The implications of this operon on the germination potential of the spores of *C. sporogenes* PA 3679 are unclear. The organisational similarity of this operon to the other ABC operons and high degree of similarity to the other ABC operons suggest it may not increase the range of germinant specificity; however it may increase the sensitivity to specific germinants. Previous studies have found that overexpression of nutrient receptors in spores can lead to increased rates of germination in response to their corresponding germinant with greater sensitivity (Cabrera-Martinez et al., 2003). The implications of variation in operon structure

upon the heat resistance of spores of *C. sporogenes* PA 3679 warrant further investigation.

While Ger receptors are the most well studied, it is likely that a number of alternative initiation mechanisms for spore germination are still to be discovered. For example, the gene *prkC*, which was identified in the genome of *C. sporogenes* PA 3679 has been associated with initiation associated with peptidoglycan fragment mediated spore germination (Shah, et al., 2008).

3.4.4 Prophage

A number of prophage were identified in the genome of *C. sporogenes* PA3679. Many phage have been investigated in *C. botulinum* predominately associated with Group III *C. botulinum* which possess BoNT encoding phage (Hwang et al., 2007; Skarin et al., 2011). The large number of prophage in *C. sporogenes* PA3679 possibly relates to the historical use of this strain in research laboratories for over 80 years. There is an unexpectedly high number of incomplete phage which are most likely associated with misassembles that could not be resolved into complete contigs during the *de novo* assembly. It was established that a minimum of 5 intact prophage were present, four identified by PHAST (Clostri_PA3679_φ3, Clostri_PA3679_φ4, Clostri_PA3679_φ5 and Clostri_PA3679_φ9) and one inferred from coverage (Clostri_PA3679_φ8). Such phage may represent a potential method for strain-specific discrimination and typing. On the basis of the complete gene BLASTN analysis it was apparent that the prophage represented the majority of strain specific DNA sequences in *C. sporogenes*.

Strong evidence was provided that Clostri_PA3679_φ8 is a temperate phage which had been induced into the lytic phase prior to extraction. Coverage of this phage was 60 times compared to the rest of the genome, indicative of a large number of copies of this phage. Typically phage induction is achieved through the addition of antibiotics such as Mitomycin C or Norfloxacin (Nale et al., 2012) and is associated with DNA damage (Kuzminov, 1999); however no specific induction of phage was attempted suggesting that this induction was spontaneous. A more in depth investigation into the ability of this phage to infect and/or lyse *C. botulinum* and other foodborne pathogens

may present future directions in producing phage mitigation strategies for the food industry.

The number of prophage demonstrates that *C. sporogenes* has the ability to acquire a significant number of phage, and thus demonstrates likely phage-mediated horizontal transfer of genetic material. Further to this, it increases the possibility that phage may have been associated with the acquisition of BoNT in Group I *C. botulinum*.

Another question generated from the identification of a large number of prophage is the effect that this may have on the niche adaption and fitness of *C. sporogenes*. It has been increasingly observed that prophage can provide benefits to their host. In the case of *C. sporogenes* PA3679 these phage may play a role in alteration of spore characteristics. It is increasingly recognised that a relationship exists between spore formation and some phage in other species either through prophage disruption of sporulation σ factors (Haraldsen and Sonenshein, 2003) or through the presence of σ factors encoded on the phage (Schuch and Fischetti, 2009). Due to the poorly described sporulation pathway in *Clostridium* it was not possible to predict whether the location of these prophage could influence the sporulation kinetics of this strain. However, differences in prophage between isolates exhibiting differential heat resistance may provide some evidence for future work in the investigation of phage mediation of spore characteristics (Chapter 6).

3.5 Concluding remarks

This study provided a draft genome assembly of *C. sporogenes* PA 3679 (NCTC 8594), an important organism in predictive food microbiology. The assembly enabled confirmation of the strong phylogenetic relationship between *C. sporogenes* PA 3679 and *C. botulinum.* The significant contribution of prophage to the total genome was identified. Further it was also demonstrated that one of these prophage (Clostri_PA3679_φ8) was a temperate phage capable of induction as evidenced by an increased coverage of sequencing data.

In regards to the heat resistant properties of *C. sporogenes* PA 3679 it was identified that this strain had an unusual number and architecture of germination operons. Whilst these cannot be definitively linked to heat resistance characteristics, it is likely

that this strain has an increased sensitivity or range of germinants. This suggests that that *C. sporogenes* PA 3679 may more easily germinate following an inimical process.

The genome of *C. sporogenes* contains 4111 genes and like many of the clostridia the functions of most of these genes are unknown. Future work may focus on identifying the function of defined gene sets. Furthermore this genome provides insight into the relationship between *C. sporogenes* PA 3679 and Group I *C. botulinum* and the acquisition of BoNT within this lineage (see Chapter 4).

Chapter 4 - Phylogenetic and pan-genomic analysis of *Clostridium sporogenes* and Group I *Clostridium botulinum*

4.1 Introduction

In chapter 3, a draft genome of *C. sporogenes* PA 3679 was presented and on the basis of 16S rRNA the phylogenetic position of the strain established. A close phylogenetic relationship between *C. sporogenes* and Group I *C. botulinum* was demonstrated, however construction of a tree based only on the 16S rRNA gene was unable to resolve significant genetic variation within the clade. Comparative analysis methods utilising multiple genes provide greater resolution and discriminatory power amongst closely related strains (Zeigler, 2003). Two such approaches are multi locus sequence typing (MLST) and pan-genomic analysis.

Amongst the most widely used molecular subtyping approaches is MLST. The method employs sequencing a number of specific housekeeping genes of interest in selected isolates. Unique sequences are assigned allelic numbers for each gene and this profile is used to assign a sequence type (ST). Multiple MLST schemes have been proposed for Group I *C. botulinum*. The scheme developed by Jacobson et al., (2008) utilised seven housekeeping genes (*oppB, hsp60, dnaE, aroE, aceK, mdh* and *recA*) to classify seventy three Group I *C. botulinum* strains into 24 ST's. Another MLST scheme by Olsen et al., (2014) utilised the genes *adh, atpH, gyrB, proC, rpoD* and *spo0A* to resolve fifty one *C. botulinum*/*C. sporogenes* strains into 37 ST groups. While such schemes provide adequate interspecies resolution they lack resolving power beyond the specific genes being interrogated.

An alternative approach to strain characterisation is to investigate the relationship of strains within the wider context of a species pan-genome. The concept of the microbial pan-genome is an increasingly popular holistic paradigm to genetically characterise a species (Mira et al., 2010). The pan-genome is the sum of the 'core' genome and 'accessory' genome. The 'core' genome consists of genes that are present in all strains whilst the 'accessory' genome is those genes present in a limited number

of strains. Pan-genomic analysis either through whole genome sequencing (WGS) or microarray methods provides significantly more insight compared to traditional methods of analysis both in regards to discriminatory power and also phenotype prediction (Medini et al., 2005; Hall et al., 2010). Previous authors have investigated the pan-genome of foodborne pathogens including *Listeria monocytogenes* (Kuenne et al., 2013), *Salmonella enterica* (Jacobsen et al., 2011), and *Escherichia coli* (Rasko et al., 2008). The pan-genome of individual strains may be quite extensive with the current size of the pan-genome of *E. coli* estimated to consist of over 16,000 genes and a core genome of only ~2200 genes (Rasko et al., 2008).

Increasingly, pan-genomic analyses have moved from description of single species (as defined by a taxonomic classification) to phylogenetical units compromising of multiple species. A combined pan-genome of *E. coli*, *S. enterica* and *Shigella* spp., (Gordienko et al., 2013) highlighted the existence of a common periphery. Similarly a pan-genome study of *Bacillus* spp. (Alcaraz et al., 2010) identified the core genome and identified the functional and evolutionary history of over 10 different *Bacillus* species.

In this chapter, the genetic relationship between toxigenic and non-toxigenic strains of *C. botulinum* is investigated using MLST and analysis of a combined pan-genome of Group I *C. botulinum* and *C. sporogenes*. The relationship of six *C. sporogenes* isolates to Group I *C. botulinum* was identified using the MLST approach of Jacobson et al., (2008), and compared to *C. botulinum* strains exhibiting varying BoNT subtypes. Following, this the pan-genome of *C. botulinum*/*C. sporogenes* was described to identify genetic biomarkers that may differentiate these organisms and explore the genetic diversity within this group.

4.2 Materials and Methods

4.2.1 Isolates

In addition to *C. sporogenes* PA 3679 (NCTC 8594) described in Chapter 3, another five isolates of *C. sporogenes* were included in the MLST analysis. Strains ATCC 19404 and ATCC 3584 were obtained from the University of New South Wales. Strains FRR B0151 and FRR B0157 were obtained from the bacterial culture collection of the Division of Animal, Food and Health Sciences, CSIRO, North Ryde, Australia. A second isolate of *C. sporogenes* PA 3679 (ATCC 7955) was obtained from the American Type Culture Collection.

4.2.2 Culture conditions and DNA extraction

Isolates were cultured and DNA extracted as described in section 3.2.1 and 3.2.2.

4.2.3 Sequencing facilities and platform

Extracted genomic DNA was sent to the Ramaciotti Centre (UNSW, Randwick, Australia) for sequencing on the MiSeq (Illumina) platform. Samples were prepared for sequencing with the Nextera XT preparation Kit (Illumina). Samples were indexed.

4.2.4 Alignment of MiSeq reads to the draft genome of *C. sporogenes* PA3679

Reads were aligned to the pseudomolecule of *C. sporogenes* PA3679 constructed in Chapter 3.0 using BWA-MEM v0.7.4 (Li, 2013). Consensus sequences for selected genes were extracted using Geneious v6.1.4.

4.2.5 Multi locus sequence typing (MLST) of *C. sporogenes*

For the purposes of multi-locus sequence typing the coding sequences for seven loci (*aroE, mdh, aceK, oppB, rpoB, recA* and *hsp*) were extracted from the consensus

sequence of each isolate. A number was assigned to each distinct locus allele according to the *C. botulinum* pubMLST database (pubmlst.org/cbotulinum).

4.2.6 Phylogenetic analysis of *C. sporogenes* and Group I *C. botulinum*

All seven coding sequences acquired from the MLST analysis were concatenated and aligned using MUSCLE (Edgar, 2004) implemented in Geneious v 6.14. A maximum likelihood tree was constructed using the one-click function in Phylogeny.fr (http://www.phylogeny.fr/version2_cgi/simple_phylogeny.cgi) (Dereeper et al.,2008).

4.2.7 Construction of the *C. botulinum/C. sporogenes* pan-genome

The *C. botulinum/C. sporogenes* pan-genome was from the genomes listed in Table 4.1 using Gview server. The seed genome was *C. botulinum* str. Hall (NC_009698). The pan-genome was constructed by querying CDS using BLASTN with a minimum alignment length of 100 bp and an strict query similarity cut-off of 80%. Regions containing phage elements in the resultant pan-genome file were identified using PHAST (Zhou et al., 2011).

Table 4.1 List of completed Group I *C. botulinum* genomes available from the NCBI reference database (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/)

Genome	Genome size* (Mbp)	Toxin serotype (Genomic location)	NCBI Accession no.	Status
***C. botulinum* Hall**	3.76	A1 (chromosome)	NC_009698	Complete
***C. botulinum* ATCC 19397**	3.86	A1 (chromosome)	NC_009697	Complete
***C. botulinum* ATCC 3502**	3.89	A1 (chromosome)	NC_009495	Complete
***C. botulinum* Langeland**	4.00	F (chromosome)	NC_009699	Complete
***C. botulinum* 230613**	3.99	F (chromosome)	NC_017297	Complete
***C. botulinum* H04402 065**	3.92	A5 (chromosome)	NC_017299	Complete
***C. botulinum* Ba4 657**	3.98	Ba (plasmid)	NC_012658	Complete
***C. botulinum* Loch Maree**	3.99	A3 (plasmid)	NC_010520	Complete
***C. botulinum* Kyoto**	4.16	A2 (plasmid)	NC_012563	Complete
***C. botulinum* Okra**	3.96	B1 (plasmid)	NC_010516	Complete
***C. sporogenes* NCTC 8594**	4.18	n/a	This Study	Draft
***C. sporogenes* ATCC 15579**	4.10	n/a	NZ_ABKW00000000	Draft

4.2.8 Calculation of the core genome, unique genes, signature genes.

The core genome was calculated by identification of CDS's that shared >80% BLASTN similarity and were present in all 12 genomes employed in the pan-genome. Unique genes were defined as CDS's that did not share >80% BLASTN similarity to CDS's in any other genome. Signature genes for *C. botulinum* were defined as CDS's that were not present in both *C. sporogenes* ATCC 15579 and *C. sporogenes* NCTC 8594; conversely the *C. sporogenes* signature genes were not present in the 10 *C. botulinum* completed genomes but were present in both *C. sporogenes* strains.

4.3 Results

4.3.1 MLST of *C. sporogenes* strains

An additional five isolates of *C. sporogenes* were selected in conjunction with the draft genome data for *C. sporogenes* NCTC 8594 (derived in Chapter 3) and *C. sporogenes* ATCC 15779. Sequences of the *aceK, aroE, hsp, mdh, oppB, recA* and *rpoB* genes were queried against the *C. botulinum* MLST database for all isolates. All *C. sporogenes* strains with the exception of strain NCTC 8594 were unambiguously assigned an allelic number for each gene. Point mutations were observed in both the *aroE* and *mdh* genes of *C. sporogenes* NCTC 8594 representing new alleles for *aroE* and *mdh.*

Allelic profiles provided unambiguous ST's for FRR B0151, FRR B0157 and ATCC 15579. The remaining strains were assigned the closest ST based on the number of matching loci.

Table 4.2 MLST profiles of *C. sporogenes* strains based on the scheme of Jacobson et al., 2008.

	NCTC 8594	ATCC 7955	FRR B0151	FRR B0157	ATCC 19404	ATCC 3584	ATCC 15579
aceK	3	29	3	3	29	29	34
aroE	2 (34:G->T)	22	2	2	2	22	26
hsp	16	16	16	16	16	15	14
mdh	4 (449:C->A)	4	4	4	4	4	22
oppB	3	26	25	25	20	26	26
recA	2	17	17	17	2	17	18
rpoB	4	4	18	18	4	4	19
Nearest match ST (matching loci)	**62 (5/7)**	**54 (6/7)**	**65 (7/7)**	**65 (7/7)**	**72 (6/7) or 83 (6/7)**	**54 (5/7)**	**70 (7/7)**

4.3.2 Phylogenetic relationship between *C. sporogenes* isolates and Group I *C. botulinum*

Phylogenetic analysis between the *C. sporogenes* isolates and *C. botulinum* was conducted using the MLST sequence data. Concatenated MLST sequences of 138 *C. botulinum* isolate sequences representing both BoNT type A and B were retrieved from the *C. botulinum* MLST database (Figure 4.1). The selected sequences represented BoNT/A isolates investigated by Jacobson et al., (2008) and Lúquez et al., (2012); and BoNT/B isolates examined by Umeda et al., (2009). Distinct differences in the sequences were observed between the *C. sporogenes* strains (Figure 4.1, rows

137-144) and the majority of *C. botulinum* strains. A tree describing the position of *C. sporogenes* within this group of *C. botulinum* was constructed (Figure 4.2). Included in the phylogenetic tree were strains that encoded BoNT on the chromosome (black text) and Type B isolates that contained plasmid-borne toxin genes (blue text).

The *C. sporogenes* isolates added in this study (NCTC 8594, 19404, FRR B0151, FRR B0157, ATCC 3584 and ATCC 7955) formed a discrete cluster. Included in this cluster was *C. botulinum* str. A207. This cluster fits within a wider clade that also included *C. sporogenes* ATCC 15579 and *C. botulinum* strains Osaka05 and 82E00751; this larger group was divergent from the majority of isolates included in the analysis. A notable exception to the primary *C. sporogenes* cluster was isolate *C. sporogenes* A4 (described by Jacobson et al., 2008) which was co-located with a number of type A *C. botulinum* strains.

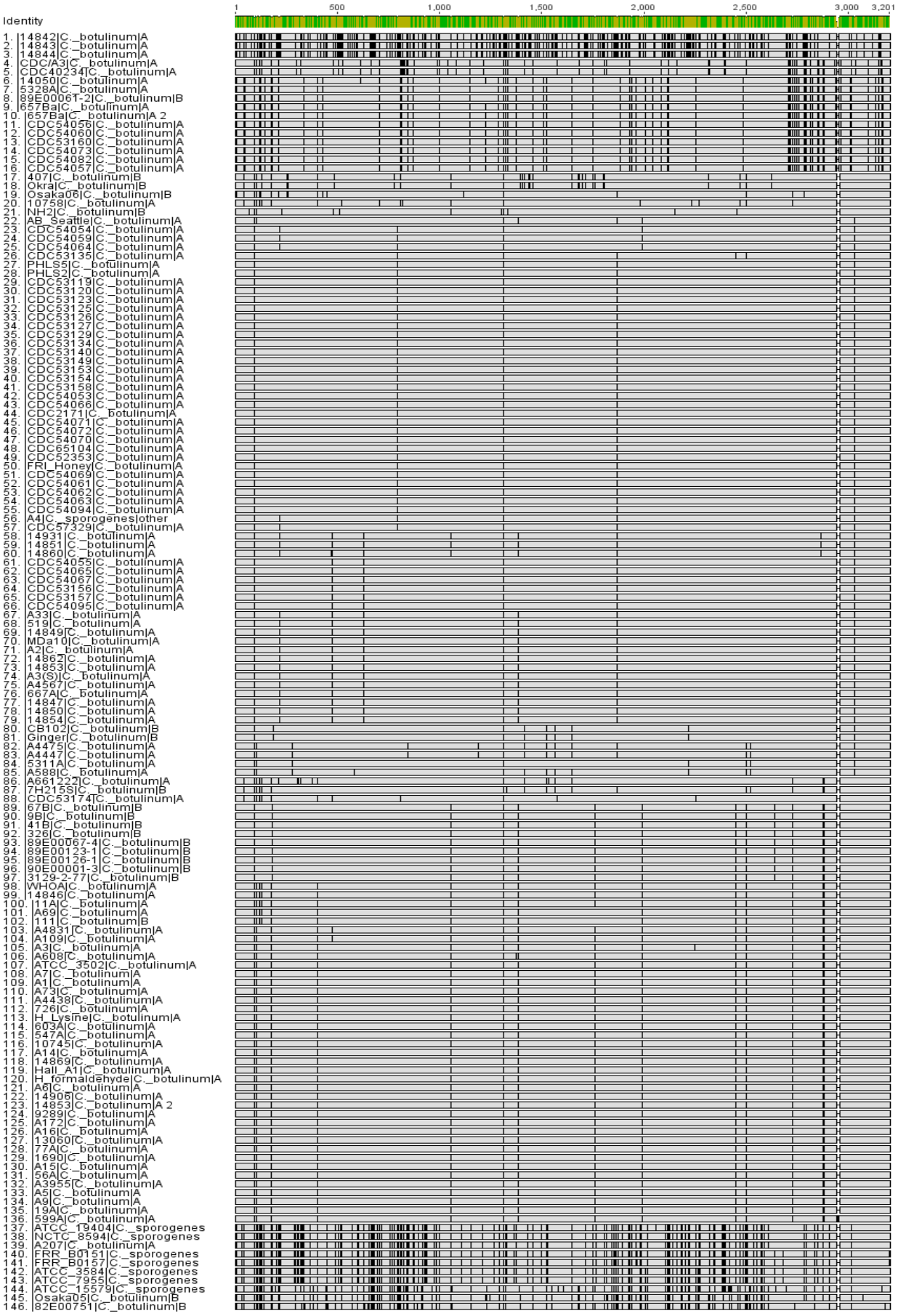

Figure 4.1 Alignment of concatenated sequences MLST sequences of *C. botulinum* obtained from http://pubmlst.org/cbotulinum/ (accessed:22/03/2013) and *C. sporogenes* strains FRR B151, FRR B157, NCTC 8594, ATCC 15579, ATCC 19404 and ATCC 3584.

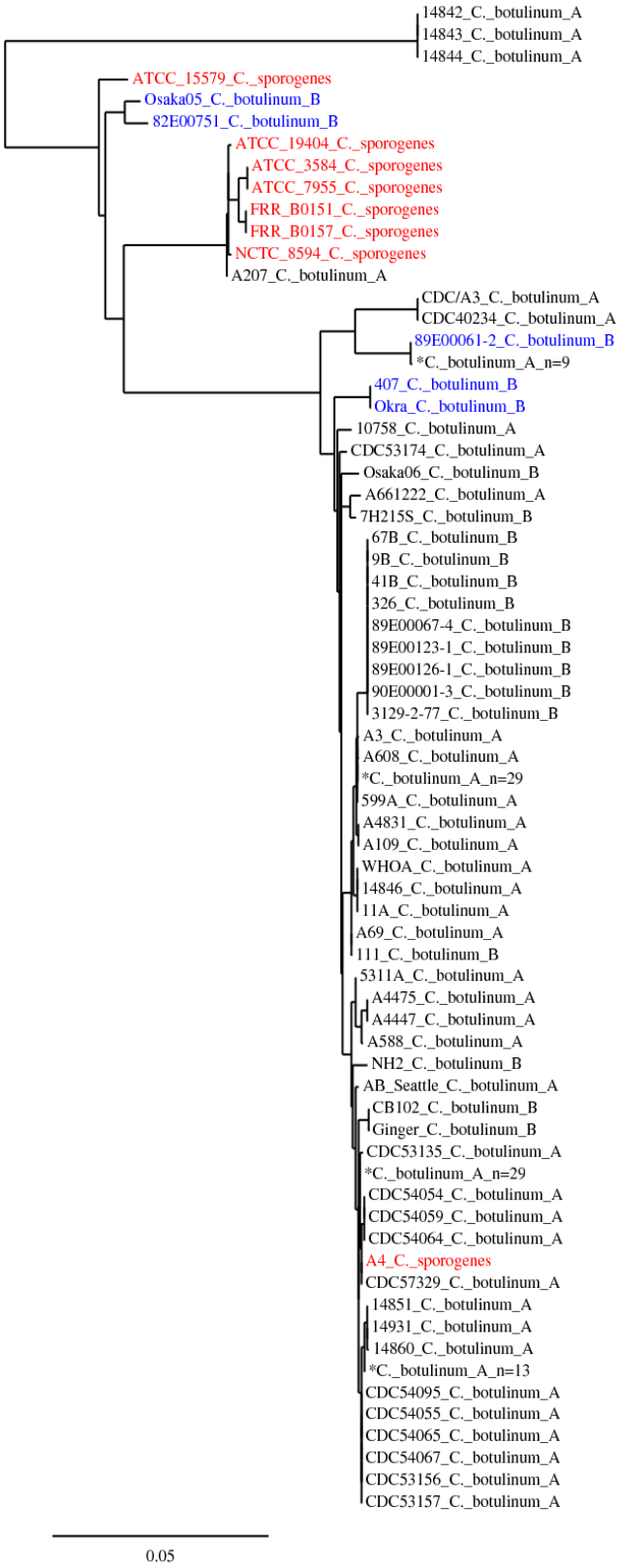

Figure 4.2 Phylogenetic maximum-likelihood tree of toxin type BoNT/A and BoNT/B Group I *C. botulinum* and *C. sporogenes* isolates. Non-highlighted strains have BoNT toxins located on the chromosome or at an unverified genomic location. Red text: *C. sporogenes*; blue text: *C. botulinum* with BoNT/B gene located on plasmid.

4.3.3 Pan-genomic analysis

Whilst the MLST study of *C. sporogenes* provided phylogenetic information it was unable to differentiate beyond the selected candidate genes. As such, a pan-genomic investigation of Group I *C. botulinum* and *C. sporogenes* was conducted.

The pan-genome was constructed by sequentially querying CDS's from each strain against the seed genome. Unique CDS's were then concatenated to the end of the seed genome before the next genome was queried against the extended seed genome. As such the configuration of pan-genome is variable dependent on the order of genome addition. The seed genome that formed the preliminary backbone was *C. botulinum* str. Hall.

A total of 12 genomes were used in the *C. botulinum*/*C. sporogenes* pan-genome. The pan-genome was constructed using ten completed Group I *C. botulinum* reference genomes and two *C. sporogenes* draft genomes, including *C. sporogenes* PA 3679 (NCTC 8594). The Group I *C. botulinum* representatives comprised 3 BoNT/A1, 1 BoNT/A2, 1 BoNT/A3, 1 BoNT/A5, 1 BoNT/Ba, 1 BoNT/B1 and 2 BoNT/F strains.

Statistics for pan-genomes constructed using either a combination of Group I *C. botulinum* and *C. sporogenes* or excluding *C. sporogenes* from the analysis are shown in Table 4.3. The combined *C. botulinum/C. sporogenes* pan-genome consists of 6,059,497 bp with a total genome of 8799 CDS's. The core genome consists of 1590 CDS's or 18.1% of the total pan-genome. (Appendix C). Of the 7209 CDS's belonging to the accessory genome approximately half were unique to single strains. In comparison the *C. botulinum* pan-genome (i.e. excluding *C. sporogenes*) consisted of a smaller number of CDS's (6943) however it had a larger core genome of 2001 CDS's.

Table 4.3 Pan-genome statistics using only Group I *C. botulinum* and a combination of Group I *C. botulinum* and *C. sporogenes*.

	Total genome CDS's	**Core genome CDS's**	**Accessory genome CDS's (strain specific)**
Group I *C. botulinum* only	6943	2001	4942 (2656)
Combined Group I *C. botulinum* and *C. sporogenes*	8799	1590	7209 (3564)

Visualisation of the pan-genome pseudomolecule is shown in Figure 4.3. The core genome is located in the first ~3.7 Mbp of the ring. Unique CDS's belonging to individual strains are located in regions after 3.7 Mbp and visualised as larger blocks. These blocks varied in size between ~200 kbp (for ATCC 3502) to ~900 kbp (for *C. sporogenes* NCTC 8594). Prophage regions were identified using PHAST in each of these blocks (Figure 4.3, outer ring). Notably in *C. botulinum* strains ATCC 3502, Ba4 and Kyoto, over half of the unique CDS's were associated with a prophage. Given the disjointed concatenation of CDS's within these blocks it is likely that some phage elements were not identified by the prophage identification tool.

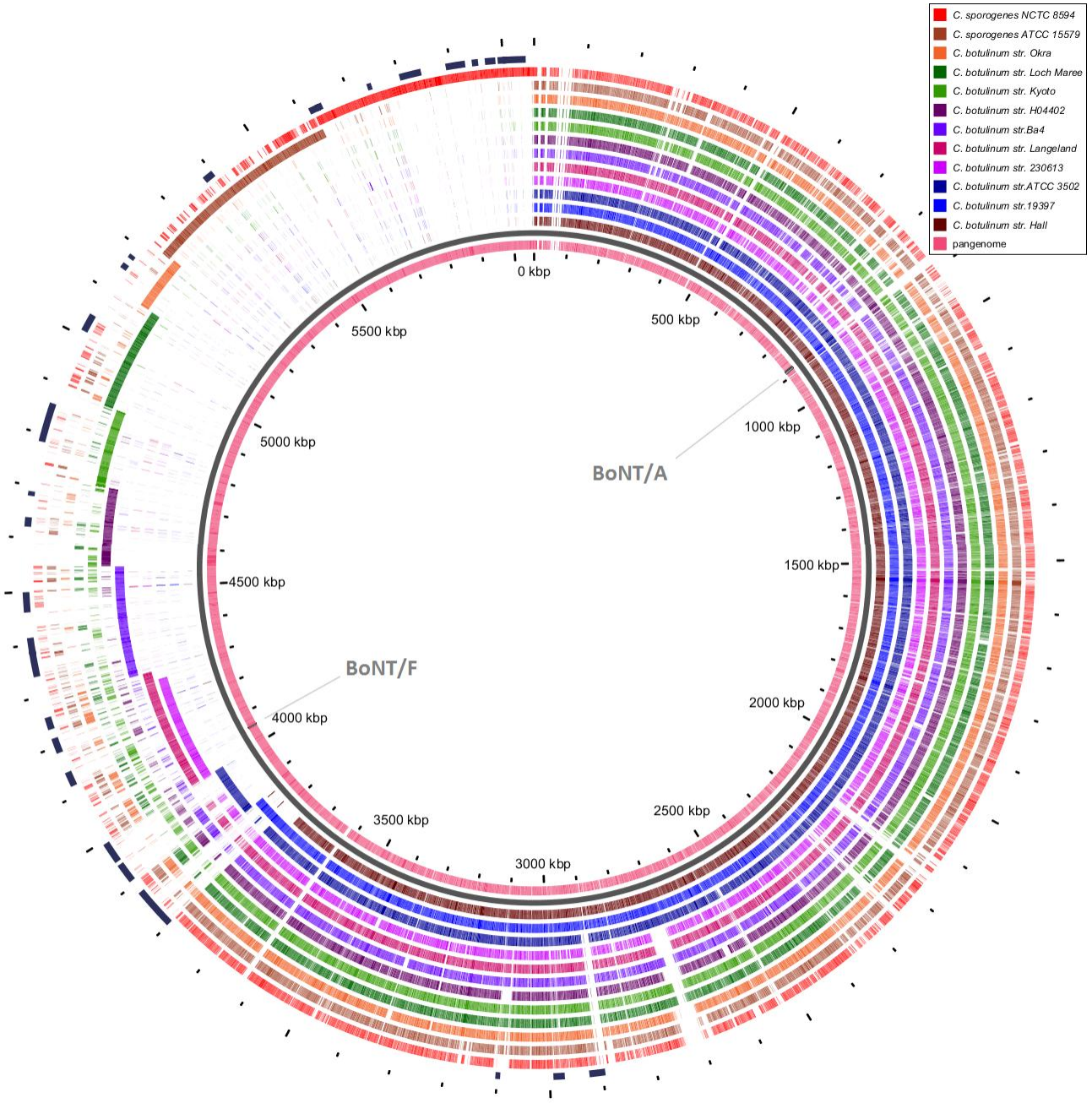

Figure 4.3 Pan-genome of *C. sporogenes* and Group I *C. botulinum*. The rings, listed from the inner to outer tracks are as follows: inner coral ring represents the constructed pan-genome. The grey inner track is a backbone for visualisation purposes. Genomes of *Clostridium* strains from inside to outside are: *C. botulinum* str. Hall (brown ring), str. 19397 (blue), str. ATCC 3502 (dark blue), str. 230613 (pink), str. Langeland (red), str. Ba4 (purple), str. H04402 065 (violet), str. Kyoto (light green), str. Loch Maree (dark green), str. Okra (orange); *C. sporogenes* str. ATCC 15579 (brown), str. NCTC 8594 (red). The outer dark blue rings are prophage regions identified using PHAST.

The effect of total genomes employed in the analysis on the number of CDS's was calculated (Figure 4.4 and Appendix D). The order of addition of genomes was as listed in Table 4.1. With increasing numbers of genomes a relatively shallow decrease in the core genome size (green) was observed. In contrast a rapid increase in the number of CDS's is seen in the accessory genome (blue) representing the significant diversity between strains and an 'open' pan-genome.

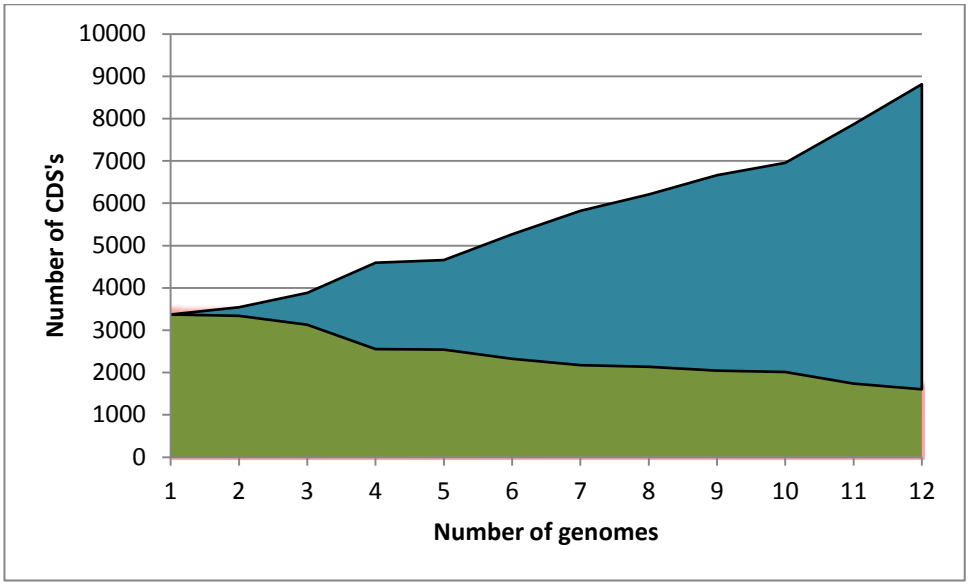

Figure 4.4 Effect on the number of genomes on the number on the core (green) and accessory (blue) genomes in the *C. botulinum*/*C. sporogenes* pan-genome.

4.3.4 Genetic differentiation between *C. sporogenes* and Group I *C. botulinum*

Genes unique to the *C. sporogenes* genomes were identified from the pan-genome (Appendix E). In total, 262 loci were identified as being unique to *C. sporogenes* and absent in the Group I *C. botulinum* genomes. Within this group 99 CDS's were identified as encoding hypothetical proteins. A number of loci associated with iron uptake and utilisation were identified including CDS encoding a ferrous iron transport protein B (IYC_09854), an iron (III) dicitrate-binding periplasmic protein (IYC_02714), a ferric enterobactin transport ATP-binding protein FebC (IYC_02724), an iron chelate ABC transport solute-binding protein (IYC_03878) and an iron chelate ABC transporter ATP-binding protein (IYC_03878). Furthermore, a number of notable genes associated with antibiotic resistance were identified,

including genes encoding a chloramphenicol acetyltransferase (IYC_18650), a permease (IYC_02694) associated with glycopeptide antibiotic resistance and a VanZ family membrane protein (IYC_14203).

A gene set of unique CDS's to the *C. botulinum* genomes was also prepared (Appendix F). Only 165 genes were identified in this set including 52 encoding hypothetical proteins. Within this set were a number of proteins with similar predicted function as those belonging to the *C. sporogenes* specific gene set due to the strict query similarity cut-off.

4.3.5 Comparison of analogous toxin regions in *C. sporogenes* and strains of *C. botulinum*

The relationship between *C. sporogenes* and *C. botulinum* was further investigated by interrogation of those genes associated with BoNT production, the differentiating phenotypic characteristic between the species. Of the ten *C. botulinum* strains employed in the study four of the strains (Hall, 19397, ATCC 3502 and H04402) contained BoNT/A subtypes encoded on their chromosome and two of the strains (Langeland F and 230613) contained BoNT/F subtypes (Figure 4.3). The remaining *C. botulinum* strains had BoNT encoded on plasmids which were not included in the pan-genome analysis.

None of the genes encoding BoNT or ANTP's were present in either strain of *C. sporogenes*. Comparison of representative genomic locations associated with the toxin regions in *C. botulinum* were made with *C. sporogenes* NCTC 8594 (Figure 4.5). The *oppB/brnQ* in *C. botulinum* str. ATCC 3502 and *arsC* operon in *C. botulinum* str. Langeland F were chosen as representative regions. These operons were selected as they are the only the two chromosomal locations in Group I *C. botulinum* where BoNT has been identified (Hill et al., 2009). It was observed that *C. sporogenes* contained putative acetyltransferase (IYC_09904), isochorismatase (IYC_09899) and hypothetical protein (IYC_09894) genes in the *oppB/brnQ* operon that were not apparent in the same region of *C. botulinum* ATCC 3502 (Figure 4.5a). In comparison to the *arsC* operon of Langeland F (Figure 4.5b), the most significant difference in *C.*

sporogenes was the presence of a gene encoding a multidrug resistance ATP transporter (IYC_10044).

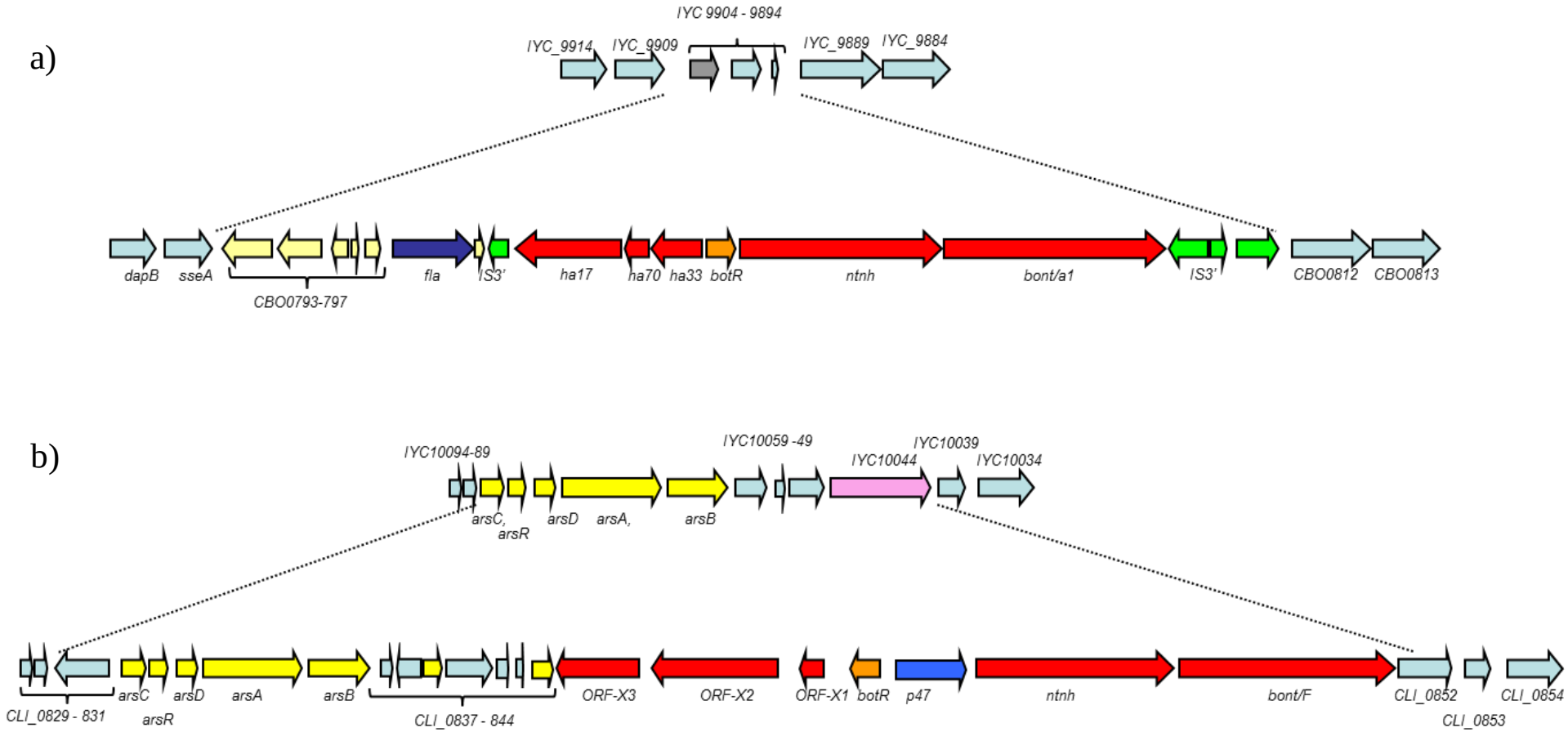

Figure 4.5 a) Comparison of *C. sporogenes* NCTC 8594 (top) and genomic region containing the toxigenic gene cluster present in *C. botulinum* str. A3502 (893870..920553) (bottom). b) Comparison of the *C. sporogenes* NCTC 8594 (top) and genomic region containing the toxigenic gene cluster present in *C. botulinum* str. Langeland F (861487..891309) (bottom).

4.4 Discussion

A number of key outcomes are described in this study: i) the phylogenetic relationship between *C. sporogenes* isolates and Group I *C. botulinum* was explored and identified divergence within *C. sporogenes* strains indicating that phylogenetic approaches are unsuitable to discriminate between these toxigenic and non-toxigenic strains; ii) those *C. botulinum* strains that clustered with non-toxigenic strains were identified to have BoNT encoding plasmids; iii) the pan-genome of *C. botulinum*/*C. sporogenes* was described, which provided high resolution into the genomic diversity displayed within the isolates tested, iv) the importance of phage in the genomic diversity of this group of organisms was highlighted by identification of unique genes and v) gene sets serving as potential biomarkers for differentiation between *C. botulinum* and *C. sporogenes* were identified.

The phylogenetic relationship between *C. sporogenes* and *C. botulinum* was assessed on the basis of MLST typing data. This study found that the *C. sporogenes* isolates investigated (with the notable exception of *C. sporogenes* str. A4) were closely related. The one strain of *C. sporogenes* that did not cluster in the group was found to reside within the same clade as a large number of BoNT/A *C. botulinum*. This finding was consistent with those of previous authors who have described the incongruence between the phylogeny and toxin types of *C. botulinum* (Collins and East, 1998; Hill et al., 2007). The presence of a non-toxigenic strain within a larger toxigenic *C. botulinum* clade indicates that the acquisition and/or loss of BoNT has occurred multiple times throughout the wider *C. sporogenes*/*C. botulinum* clade. Furthermore, it confirms that phylogeny is not necessarily associated with toxigenicity. As such phylogenetic methods employing single genes or even MLST methods are incapable of providing adequate resolution to determine toxigenicity of strains.

Three *C. botulinum* strains were assigned to the same *C. sporogenes* cluster. Of these two strains Osaka05 and 82E00751 contained BoNT/B located on plasmids. The third strain was *C. botulinum* A207 which produces BoNT/A. The genomes of these strains were unavailable preventing further analysis of this finding. The observation of clustering of plasmid-borne BoNT/B strains of *C. botulinum* with *C. sporogenes* provides further evidence that horizontal gene transfer is associated with the acquisition of BoNT in this group of organisms. Conjugative plasmid mediated

transfer of BoNT encoding plasmids has been observed in *C. botulinum* (Marshall et al., 2010) suggesting that this is a likely mechanism for the dissemination of BoNT encoding genes throughout this group of organisms. While the genomic location of BoNT in *C. botulinum* A207 is not known, one may predict that toxigenicity in this strain is likely associated with plasmids. Alternatively, this strain may represent a distinct lineage from the more common Type A *C. botulinum* strains.

A limitation of the phylogenetic analysis presented is the selection bias towards toxigenic strains within the database. The MLST database employed in this study contained clinically relevant isolates rather than environmental isolates. Unsurprisingly this means there were significantly more *C. botulinum* strains than *C. sporogenes* strains. This is in contrast to previous surveys which have found ratios of 10,000:1 *C. sporogenes* to *C. botulinum* spores in soils (Hauschild and Dodds., 1993). A cursory view of the phylogenetic tree constructed may suggest that it is the *C. sporogenes* strains that are the outliers from a larger distribution of *C. botulinum* type organisms; however it is just as likely that in the environment it is the *C. botulinum* strains which represent the distinct lineage and are outliers. With the growing amount of environmental metagenomic data being produced, it soon may be possible to draw more specific conclusions about the abundance of non-toxigenic and toxigenic strains in the environment. A survey of environmental isolates may provide evidence for a whether selective environmental pressures assist in the maintenance or elimination of genes involved in toxin production.

The construction of the pan-genome of Group I *C. botulinum*/*C. sporogenes* provided insight into the genomic diversity exhibited by *C. sporogenes* and Group I *C. botulinum*. The level of diversity observed was surprisingly large given the small sample size of genomes (n=12). As more genomes were added to the pan-genome an increase in the size of the accessory genome was consistently observed. This would classify the pan-genome of *C. sporogenes*/*C. botulinum* an 'open' genome. Pan-genomes that are 'open' represent strains that actively promote the exchange of genetic material. As both a soil and human intestinal bacterium it is unsurprising that such high levels of genetic exchange are observed in *C. sporogenes*, as exposure to a diverse range of environmental conditions requires high levels of adaptability (Tettelin et al., 2008).

The significant role of phage in the genetic diversity within *C. botulinum* and *C. sporogenes* was highlighted in the pan-genomic analysis. The identification of putative prophage regions in the strain specific gene sets depicts that these mobile elements are a major source of observed variation between strains within the *C. botulinum*/*C. sporogenes* clade.

The identification of 262 unique genes present in *C. sporogenes* with no orthologs (at an 80% identity level) in *C. botulinum* presents a future opportunity to discover potential biomarkers for toxigenicity. However, as a significant proportion of these genes were associated with phage and/or encoded hypothetical proteins, it is unclear whether these genes are truly specific to *C. sporogenes* and this therefore remains to be further validated in a wider range of *C. sporogenes* strains. Furthermore, these biomarkers must be tested against an increased range of *C. botulinum* including those with plasmid encoded BoNT.

4.5 Concluding remarks

Phylogenetic analysis of *C. sporogenes* identified that the majority of *C. sporogenes* isolates displayed a high degree of similarity to each other and Group I *C. botulinum*. Clades containing both toxigenic and non-toxigenic strains were identified indicating that phylogenetic approaches are unsuitable for discrimination of the two species. Approaches such as the identification of *C. sporogenes* strain specific genes may provide greater opportunity to discriminate between non-toxigenic and toxigenic strains; however evidence for the lateral acquisition of BoNT suggests that this approach may not be feasible. The clustering of *C. botulinum* strains containing plasmid-encoded BoNT genes within the primary *C. sporogenes* clade is indicative that *C. sporogenes* strains are capable of the acquisition of BoNT.

The pan-genome of *C. sporogenes*/Group I *C. botulinum* was shown to be 'open' with over ~82% of the pan-genome belonging to the accessory genome, highlighting significant genetic diversity across the group. This diversity between strains was heavily shaped by the integration of prophage of the genomes included in the analysis.

Increasing both the number and geographical spread of isolates would significantly improve the resolution and conclusions of this study. Future directions could include

greater emphasis on a broader range of isolates, both non-toxigenic and toxigenic to identify the evolution of BoNT within this group of organisms. Further investigation into the nature of phage in *Clostridium* spp. may additionally provide insights into the genetic diversity of *C. sporogenes* and *C. botulinum*. Both of which could aid in the development of novel risk management strategies against Group I *C. botulinum* in the food chain.

Chapter 5 - Flow cytometric analysis of variation in the structure of individual spores of *Clostridium sporogenes* PA 3679 under thermal stress

5.1 Introduction

Bacterial spores, such as those formed by *Clostridium* spp. and *Bacillus* spp. are a ubiquitous problem for the food industry. Dormant spores are extremely resistant to environmental stresses and capable of surviving preservation processes that would inactivate vegetative cells. Ensuring the safety and stability of food products in relation to spores is often achieved through the use of thermal processes, careful product formulation, or a combination of thermal treatments and additional microbial hurdles that prevent germination and outgrowth of the target spore. Those hurdles may be intrinsically associated with the product or may be introduced by design.

Clostridium sporogenes PA 3679 is a commonly employed non-toxigenic surrogate for Group I *Clostridium botulinum* in food challenge studies, particularly with regards to thermally treated products. A distinguishing characteristic of this strain is the reported higher heat resistance compared to *C. botulinum* (Diao et al., 2014), however while numerous studies have indirectly described the intrinsic heat resistant properties of *C. sporogenes* (see Review by Brown et al. 2012) very few have described these changes via direct measurement of individual spores. In order to investigate this aspect of *C. sporogenes* PA 3679 a flow cytometric approach was developed to characterise the structural variation of individual *C. sporogenes* spores during an isothermal process and a subsequent post-heating recovery period. These approaches were then applied to investigate the effects of a secondary environmental stress, NaCl, on the heterogeneity of individual spores.

An accurate assessment of the kinetic response of spores to heat treatments is a challenging aspect in the development of predictive models for spore inactivation. Traditionally, the study of the response of spores to heat inactivation is based on indirect measurements of recoverable spores. In practice, this is often

achieved by measurement of 'survivors' from a known starting concentration of spores after an applied process in a recovery medium. The use of survivor data such as this obfuscates the stochastic response of individual spores to a thermal process and thus may mask the biovariability existing within a spore population by labelling spores as either "dead" or "live". It may also mask secondary effects of the heat treatment of spores such as sublethal injury exhibiting itself through altered nutritional requirements and/or increased sensitivity to inhibitors in culture medium (Adams, 1978; Reddy et al., 2010).

Flow cytometry (FCM) is a rapid, culture-independent technique capable of providing multi-parametric measurements of single cells (Davey and Kell, 1996; Veal et al., 2000). Cells are suspended in a hydrodynamically focused stream and passed through a laser and detector system to quantify fluorescent and light emission characteristics. FCM has been increasingly applied in food microbiology, particularly for monitoring, detection and quantification of vegetative cells, where it may produce more sensitive and faster results than plate culturing techniques (Budde and Rasch, 2001).

FCM in microbial studies is commonly used in combination with specific fluorescence probes which label cells according to their physiological state (Sträuber and Müller, 2010) The fluorogenic nucleic acid dyes SYTO 9 and propidium iodide (PI) are two such probes, which upon intercalating with DNA or RNA have significantly enhanced fluorescence signals (Stocks, 2004). SYTO 9 and PI are components of the *BacLight* LIVE/DEAD kit employed as an assay to differentiate between viable and non-viable vegetative bacterial cells. Whether permeability to PI accurately represents non-viability is highly dependent on the conditions and physiological state of the cells (Shi et al, 2008; Davey and Hexley, 2011). More accurately this assay can be described as a measurement of membrane integrity based on selective exclusion of dyes. Cells with an intact membrane will permit the ingress of only SYTO 9 to nucleic acid within the cell, while cells with compromised membrane integrity allow PI to enter the cell and preferentially bind to the nucleic acid (Stocks, 2004). More recently those fluorophores have been employed to study spore germination, sporulation-related sub-population transitions and spore inactivation (Laflamme et al., 2004; Coleman et al., 2007; Cronin and Wilkinson, 2007; Mathys et al., 2007; Jones et al., 2008, Tracy et al., 2008; Kong et al., 2010;

Ferencko and Rotman, 2010, van Melis et al., 2011). These studies have all utilised species other than *C. sporogenes* or *C. botulinum* and very few have used FCM to provide quantitative measurements regarding the relative fluorescence emission of the dyes, relying on qualitative microscopic counts.

In this study FCM was applied to quantify the SYTO 9 and PI fluorescence amongst individual spores of *C. sporogenes* PA 3679 (NCTC 8594). On the basis of the dye emission characteristics sub-populations exhibiting similar structural alterations were described. In addition the impact of varying NaCl concentration (as an example of a common antimicrobial ingredient often removed for the reformulation of healthier products), both in the heating and recovery medium, on the number of spores within each of these sub-populations was measured.

5.2 Materials and Methods

5.2.1 Bacterial strain and culture conditions

C. sporogenes PA 3679 (NCTC 8594) was obtained from the National Collection of Type Cultures. All incubations were performed in an atmosphere consisting of 97% (v/v) nitrogen and 3% (v/v) hydrogen in an anaerobic workstation (Coy Laboratory Products, Michigan, USA).

5.2.2 Spore production

Spore crops were prepared in a biphasic system to enhance the spore yield compared to that achieved using a liquid medium. The biphasic medium was prepared in 650 mL tissue culture flasks (Greiner Bio-One GmbH, Frickenhausen, Germany) and consisted of 100 g Anellis base agar, which upon cooling was overlaid with 50 mL tryptone-peptone-glucose starch (TPGS) broth.

The inoculum was prepared from an initial loopful of -80°C glycerol stock culture inoculated into 10 mL of Cooked Meat Medium (CMM) and incubated for 48 h at 37°C. Subsequently, a 100 μL aliquot of the CMM culture was subcultured to 10 mL TPGS and incubated for 24 h at 37°C. A 100 μL aliquot of the TPGS culture was then used to inoculate the biphasic system.

Cultures were incubated at 37°C for 5-7 d with the suspensions examined microscopically. When the culture comprised >70% phase bright spores, the spores were harvested by centrifugation (12,000 *g* for 10 min at 5°C) and purified as described below.

5.2.3 Spore purification

Spore crops were purified by density gradient centrifugation using Urografin® (Bayer Schering Pharma AG, Berlin, Germany) as described by Nicholson and Setlow (1990) with some modifications. Pellets (which contained spores, vegetative cells and cellular debris) were suspended into 2 mL 10% (w/v) Urografin®. This suspension was gently layered onto 20 mL volumes of 70% (w/v) Urografin ® in a 50 mL centrifuge tube, and centrifuged at 8000 *g* for 1 h in a swinging bucket rotor at room temperature.

The resulting supernatant containing phase dark spores, vegetative cells and debris was discarded and the resultant pellet of phase bright spores was collected by resuspension in a small volume of chilled sterile distilled water (SDW). The spore suspension was centrifuged (12,000 *g* for 10 min at 5°C) and washed a further three times in SDW to further purify the phase bright spores and remove any residual Urografin ®. The spore crop was diluted in chilled SDW to a concentration of ~10^8 spores mL^{-1}.

The concentration and purity of the resultant crops was assessed by haemocytometer counts using phase contrast microscopy. All crops were confirmed to contain >99% phase bright spores.

5.2.4 Thermal treatments of *C. sporogenes* spores

Spores were heated in a minimal nutrient medium prepared by making a 1:10 dilution of *C. sporogenes* recovery broth (CSRB). CSRB is comprised of 5 g L^{-1} tryptone, 5 g L^{-1} peptone, 2 g L^{-1} yeast extract, 2 g L^{-1} glucose, 2.5 g L^{-1} NaCl, 2 g L^{-1} K_2HPO_4, 10 mM L-alanine, 10 mM D,L-lactic acid (sodium salt) (Sigma-Aldrich, Castle Hill, Australia), 0.5 g L^{-1} sodium thioglycollate and 1 g L^{-1} $NaHCO_3$. Three heating menstruum were employed; 1:10 CSRB with no added NaCl (HM +0%) and 1:10 CSRB supplemented with 2% and 4% (w/v) of NaCl (HM +2% and HM +4%).

Volumes of 9.9 mL of the heating menstruum were dispensed into screw-capped glass test tubes and placed in an oil bath to equilibrate to 96.0±0.2ºC. Tubes were located in the oil bath ensuring that the heating menstruum was located below the surface of the oil bath A 100 μL aliquot of the defrosted spore suspension (1.1 x 10^8 spores mL^{-1}) was injected midway into the menstruum contained in each tube.

Spores were heat treated for 6 min, 24 min and 54 min representing a mild, moderate or severe treatment respectively. Samples were immediately cooled to room temperature in a mixture of crushed ice and water. All experiments were conducted in triplicate.

Trial experiments indicated that the injection of 100 μL inoculum decreased the heating menstruum temperature by no more than 0.5°C and the come-up time to

return to a temperature ±0.2°C was within 2 minutes. The cool-down rate ranged between 8.0°C and 10.0°C s^{-1} in the ice slurry (data not shown).

5.2.5 Enumeration of recoverable spores and calculation of D-values

Recoverable spores were enumerated on CSRA pour plates (CSR broth + 1.5% (w/v) agar (Oxoid) following the preparation of serial dilutions in 0.1% (w/v) peptone water (Oxoid) from heat-treated spore suspensions. Plates were incubated anaerobically at 30ºC for 7 days. All plates were prepared in duplicate.

Survival curves were plotted as $\log_{10}$ cfu mL^{-1} versus time. The decimal reduction time at 96°C (D_{96}) was calculated from the negative reciprocal of the slopes of the regression lines, using the linear portion of the survival curves.

5.2.6 FCM observations of thermally injured *C. sporogenes* spores in a recovery medium

Heat-treated spore suspensions were suspended into 4.5 mL volumes of recovery medium consisting of either CSRB (RM +0%) or CSRB supplemented with 2% (w/v) NaCl (RM+2%). Broths were supplemented with 50 μg mL^{-1} ampicillin (Sigma-Aldrich) to prevent the proliferation of vegetative cells during the recovery period. Cultures were incubated anaerobically for 72 h at 30ºC. Samples were withdrawn and analysed after 0, 6, 12, 24, 48 and 72 h and stained as described below.

5.2.7 Staining

Spores were stained with the nucleic acid binding fluorescent dyes, SYTO 9 and propidium iodide (PI). A stock solution of SYTO 9 in water was prepared from the LIVE/DEAD® BacLight™ Bacterial Viability Kit applicator sets (L13152, Life Technologies, Carlsbad, US). SYTO 9 in a powdered form was reconstituted in 500 μL of sterile distilled water followed by filtering through a 0.2 μm syringe filter. This was stored frozen at -18°C and protected from light and used within a month of preparation. PI was kept as a stock solution of 1 mg mL^{-1} of PI in water (Life Technologies) and stored refrigerated.

Diluted spore samples were stained by mixing a 485 μL aliquot with 5 μL SYTO 9 (120 μM) and 10 μL of PI (~3 mM); resulting in final concentrations of 1.2 μM and ~30 μM of SYTO and PI, respectively. Samples were stained under anaerobic conditions at 30°C in the dark for 8 minutes prior to analysis.

5.2.8 Flow cytometry

Flow cytometric analyses were conducted on a FACSCalibur flow cytometer (BD Biosciences, San Jose, USA), equipped with a 15 mW air-cooled 488 nm argon-ion laser. Osmosol (Lab Aids, Sydney, Australia) was used as the sheath fluid. Green fluorescence (FL1-H) was collected through a 530 nm band-pass filter and red fluorescence (FL3-H) through a 650 nm long-pass filter. A minimum of 20000 events were collected on the basis of their forward-angle light scatter (FSC) and side scatter (SSC) properties. The instrument settings used throughout the experiments were FSC=E01, SSC=417 mV, FL1=602 mV, FL3=711 mV using logarithmic amplifiers. Samples were run at the low flow rate setting (12 μL/s). Online data acquisition was made with BD CellQuest Pro software (BD Biosciences) prior to offline analysis.

5.2.9 Data and statistical analysis

FlowJo version 7.5.5 (TreeStar, Ashland, USA) was used to visualise and analyse FCM data offline. Data were visualised as bivariate plots. Initially, non-spore-related events were removed by gating on the basis of FSC-H and FL1-H, followed by manually setting of quadrant coordinates set on the basis of FL1-H and FL3-H fluorescence (Figure 5.1). Non-spore related events (debris) were excluded as those events with non-specific staining characteristics (i.e. a correlation between FSC-H and FL1-H). Crossover in fluorescence between populations present in Q2 and Q3 occurred after more severe thermal treatments, in these cases quadrant coordinates were set in the red fluorescence channel at the apex formed by the two diagonal populations.

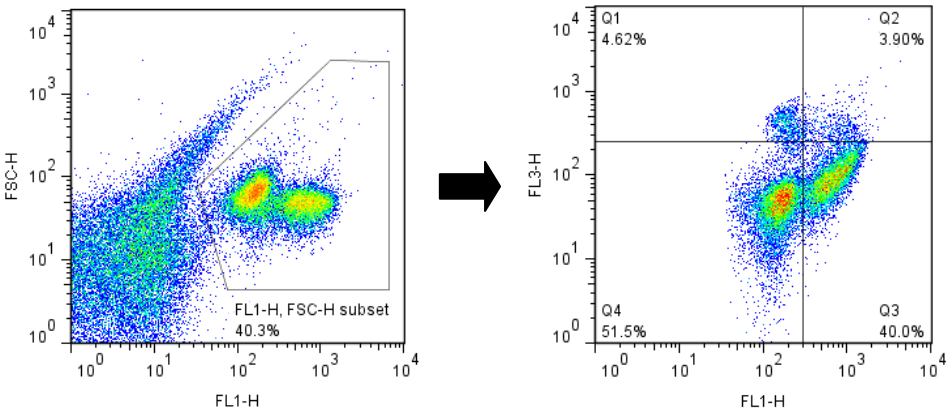

Figure 5.1 Example of the two-step FCM gating approach on scatterplots, spores were initially manually gated on the FSC-H and FL1-H channels to remove debris followed by bivariate gating into quadrants on the FL1-H and FL3-H channels.

For populations in RM +2% it was necessary to adjust the gating strategy to compensate for the refractive index mismatch with the sheath fluid and the decreased binding affinity of nucleic acid stains under ionic conditions (Marie et al., 1996). In this instance quadrant coordinates were identified as those yielding an equivalent (+/- 2%) proportion of spore-related events in each of the four quadrants as those without NaCl at t=0.

All statistical tests were calculated using R, version 2.1.40 (R Foundation for Statistical Computing, Vienna, Austria). The effect of changing NaCl concentrations within each quadrant and treatment duration were analysed with a one-way ANOVA, followed by post-hoc pair-wise comparisons of the means by Tukey's HSD. A p-value ≤ 0.05 was considered statistically significant.

5.3 Results

5.3.1 Comparative measurement of green (FL1-H), red (FL3-H) and side scatter (SSC-H) in dormant, germinating and heat-inactivated *C. sporogenes* spores

Preparations of dormant, germinating and heat-inactivated spores were used as controls to identify the emission characteristics of spores with known structural variations. The median FL1-H (green), FL3-H (red) and SSC-H (side scatter) profile of each of these preparations is shown in Table 5.1.

Table 5.1 Median channel intensity of dormant, germinating and heat inactivated spore preparations.

	Median FL1-H	Median FL3-H	Median SSC-H
Dormant	11.0 ±2.3	32.5 ±2.0	18.92 ±0.3
Germinating	44.1 ±0.3	55.2 ±0.7	17.91 ±0.4
Heat-treated	60.7 ±3.4	94.0 ±7.7	8.01 ±0.2

The preparation of dormant spores exhibited the lowest median intensity in both green and red fluorescence, indicating that both nucleic acid stains were unable to penetrate to the spore core. Dormant and germinating spores exhibited significantly higher side scatter characteristics compared to heat-injured spores.

Spores placed into a germination medium for 3 h split into two distinctive populations of dormant and germinating spores. Germinating spores had significantly higher median green fluorescence intensity and an increased shift in median red fluorescence intensity compared to dormant spores. The germinating spores and dormant spores however shared similar light scatter characteristics.

Heat inactivated spores prepared by heating at 96°C for 2 h comprised of a single population. This treatment resulted in a single population that exhibited the highest fluorescence intensity in both fluorescence channels and the lowest side scatter characteristics of all three preparations.

5.3.2 Analysis of fluorescence characteristics of individual spores in a heat-treated population

Heating purified crops of dormant spores promoted a variable response of individual spores. This was exhibited by the formation of multiple sub-populations from an original homogeneous population of dimly fluorescent SYTO 9 and PI spores. Post-treatment individual spores were observed to reside in one of three sub-populations exhibiting differing SYTO 9 and PI fluorescence emissions. These sub-populations constituted a dimly fluorescent SYTO 9 and PI population (SYTO 9/PI^{dim}), a SYTO 9 positive and PI dim population (SYTO 9^{+}) and a PI positive population (PI^{+}) (Figure 5.2). The properties and predicted structural characteristics of these sub-populations are tabulated in Table 5.2.

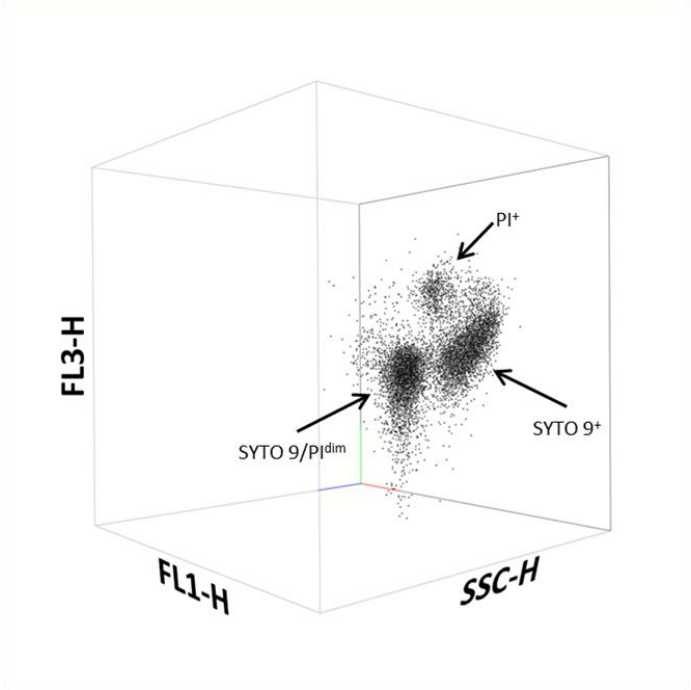

Figure 5.2 Three-dimensional scatterplot of primary sub-populations of spores identified after heat treatment at 6 minutes at 96°C.

Table 5.2 Predicted state of sub-populations of dual-stained *C. sporogenes* spores

Sub-population	Relative fluorescence intensity (SYTO 9/PI)	Relative SSC-H	Spores present in sub-population	Predicted structural characteristics
SYTO 9/PI^{dim}	Low/Low	High	Dormant spores	Spores that retain a dehydrated core.
SYTO 9^{+}	High/Low	Low	Germinating and heat injured spores	Spores with an intact but semi-permeable inner membrane and rehydrated core.
PI^{+}	Low/High and High/High	Low	Heat injured spores	Spores with a rehydrated core and compromised inner membrane integrity.

Table 5.3 Variation in logarithmic reduction and sub-populations after thermal treatments (mean±SD)

Treatment	Log reduction (CFU mL^{-1})	% SYTO 9/PI^{dim}	% SYTO 9^{+}	%PI^{+}
0 min (untreated)	n/a	99.9	<0.1	<0.1
6 min (mild)	1.75 ±0.26	55.6 ±3.1	35.1 ±2.4	8.3 ±0.7
24 min (moderate)	4.01 ±0.11	31.7 ±2.9	49.2 ±7.5	19.1 ±5.2
54 min (severe)	4.94 ±0.23	13.4 ±1.6	70.1 ±5.0	15.9 ±3.4

The percentage of spores falling into each spore sub-population was monitored after a mild, moderate and severe isothermal heat treatment at 96°C (Table 5.3). A comparison between colony forming units and the percentage of spores in each population was also undertaken to investigate the possibility of FCM as a rapid indicator of culturability.

Spores subjected to the mild heat treatment predominately remained in the SYTO 9/PI^{dim} population; however a significant proportion of spores transitioned to the SYTO 9^{+} population and a smaller percentage to the PI^{+} population.

Increases in heating period led to further reductions in the percentage of spores in the SYTO 9/PI^{dim} sub-population. A significant increase in the proportion of spores in both SYTO 9^{+} and PI^{+} sub-population was observed after the moderate heat treatment compared to the mild heat treatment. Further extension of the treatment period

resulted in more than 70% of spores in the SYTO 9^+ population, however no significant change was observed in the percentage of PI^+ spores compared to the moderate heat treatment.

5.3.4 Effect of addition of NaCl on the survival of spores during a heat process

The addition of NaCl to the heating menstruum altered the ratio of spores in the observed sub-populations in a manner dependent on the severity of the heat treatment (Figure 5.3 and Figure 5.4).

Heating in HM +2% and HM +4% resulted in increased percentages of SYTO $9/PI^{dim}$ spores after both the mild (Figure 5.4i) and moderate (Figure 5.4ii) treatments compared to those spores heated in HM +0%. In the case of mild heat treatment, this was mirrored by a coincident reduction of both SYTO 9^+ and PI^+ spores. In the case of the moderate heat treatments the increased percentage of SYTO $9/PI^{dim}$ spores was predominately reflected solely in a decrease in the percentage SYTO 9^+ spores.

Increasing NaCl concentrations did not significantly alter the spores remaining in the SYTO $9/PI^{dim}$ state after a severe heat treatment (Figure 5.4iii). It did however significantly increase the percentage of PI^+ spores with a corresponding decrease in the percentage of SYTO 9^+ spores.

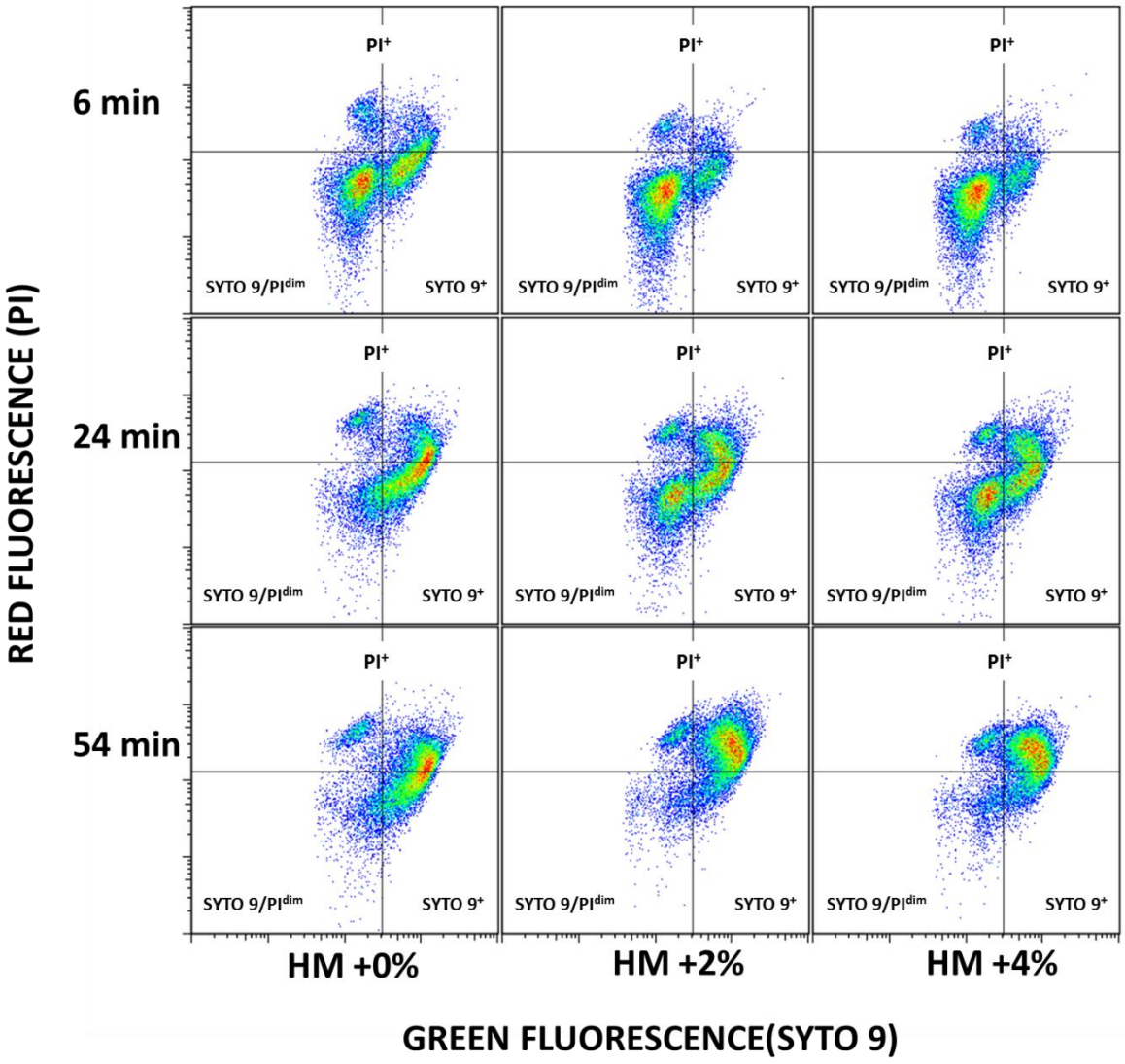

Figure 5.3 Example FCM derived bivariate plots of dual-stained spores of *Clostridium sporogenes* PA3679 after thermal treatment (96ºC±0.2) in 1:10 CSR broth at one of NaCl concentration (+0%, +2% and +4%) for 6, 24, and 54 minutes.

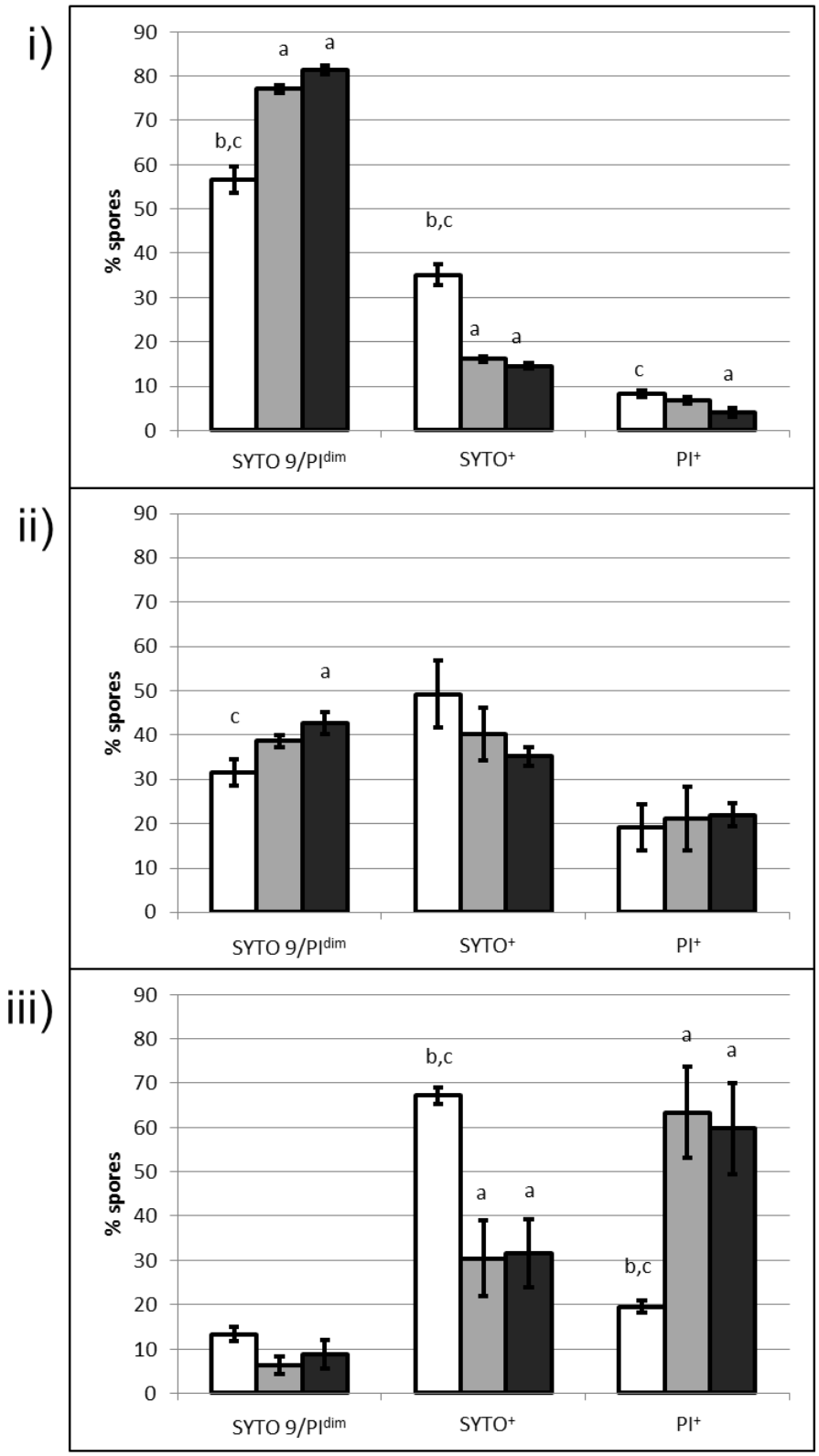

Figure 5.4 Mean percentage of events in each sub-population directly after heating spores for 6(i), 24(ii) and 54(iii) minutes in a heating menstruum containing additional NaCl concentrations of 0% (HM +0%; white), 2% (HM +2%; grey), and 4% (HM +4%; black). Error bars are standard error of the mean (n=3). Letters represent a statistical difference ($p<0.05$) within each sub-population to different NaCl concentrations (a=HM +0%, b=HM +2% and c=HM +4%).

5.3.5 Spore structural changes post heat treatment under outgrowth conditions (+/- NaCl)

Post-heating spores were resuspended in a germinant rich recovery medium in the absence and presence of additional NaCl (Tables 5.4-5.6). Spores were tracked for 72 h with the majority of spores exhibiting significant changes in their permeability to SYTO 9 and PI during this period.

The SYTO 9/PI^{dim} population of spores remained present throughout the recovery period irrespective of the heat treatment. An increase in this population was observed under certain recovery conditions (denoted by asterisk in Table 5.4 and Table 5.5) and was attributable to an additional spore sub-population within the SYTO 9/PI^{dim} population. These spores were designated as SYTO 9/$PI^{dim(\beta)}$. This SYTO 9/$PI^{dim(\beta)}$ sub-population increased the total percentage of SYTO 9/PI^{dim} population typically between 12-24 h and could not be unambiguously distinguished from SYTO 9/PI^{dim} spores on the basis of green and red fluorescence alone, however exhibited differential light scattering characteristics. Microscopic examination was inconclusive as to the identity of this sub-population.

Spores that were given the mild heat treatment in the unadulterated heating menstruum and recovery medium (HM +0%/RM +0%) were observed to have a low percentage (<10%) of PI^{+} spores at t=0 however this percentage rapidly increased within 6 h to the point that this population represented the majority of spores present for the remainder of the 72 h recovery period. A consistent decrease in the SYTO 9^{+} population was observed throughout this incubation period. In all other spore populations treated mildly a rapid increase in SYTO 9^{+} spores was observed and peaked after 12 h before declining with these spores subsequently transitioning into either a PI^{+} or a SYTO 9/$PI^{dim(\beta)}$ state. A similar population profile was observed for spores that received a moderate heat treatment (Table 5.5).

Severely heat treated spores predominately resided in the SYTO 9^{+} or PI^{+} populations throughout the recovery period. Spores heat treated in HM +2% started with a significantly higher proportion of PI^{+} than spores treated in HM +0%, however, after a 72 h recovery period the percentage of PI^{+} spores reached similar levels irrespective of the heating or recovery medium.

The majority of spores exhibited a similar transition path as those observed during a heat treatment, with the majority of spores transitioning from SYTO 9/PI^{dim} → SYTO 9^{+} → PI^{+} and/or SYTO 9/$PI^{dim(\beta)}$ over the recovery period, however the timing of transition did vary. A consistent trend was observed of spores suspended in RM +2% retaining a higher proportion of spores located in SYTO 9^{+} compared to spores suspended in RM +0%, this trend was most pronounced after 12 h, indicative that only the transition out of SYTO 9^{+} was affected by the presence of increasing NaCl concentrations.

Table 5.4 Percentage of spores (%) in each sub-population after a mild (96ºC for 6 min) thermal treatment

NaCl Concentration (Heating/Recovery)	Sub-population	time (h)					
		0	**6**	**12**	**24**	**48**	**72**
HM +0%/RM +0%	SYTO 9/PI^{dim}	55.6 ±3.1	14.1 ±4.9	20.0 ±10.0*	14.5 ±5.0*	12.1 ±2.1*	10.3 ±0.9*
	SYTO 9^{+}	35.1 ±2.4	31.0 ±2.5	30.0 ±3.7	18.3 ±1.3	13.8 ±0.6	11.1 ±2.4
	PI^{+}	8.3 ±0.7	54.9 ±5.5	50.1 ±13.8	67.2 ±5.3	74.1 ±2.7	78.7 ±3.1
HM +0%/RM +2%	SYTO 9/PI^{dim}	55.5 ±3.1	24.7 ±8.1	21.5 ±1.7	11.1 ±2.1	9.3 ±0.9	7.9 ±2.5
	SYTO 9^{+}	35.6 ±2.3	39.8 ±2.6	55.9 ±3.9	41.7 ±3.9	30.0 ±4.2	35.4 ±4.5
	PI^{+}	5.9 ±0.9	35.5 ±7.7	21.6 ±3.0	47.2 ±5.0	60.7 ±3.7	55.6 ±2.2
HM +2%/RM +0%	SYTO 9/PI^{dim}	77.1 ±0.9	15.5 ±8.0	19.9 ±0.5*	22.3 ±4.5*	28.0 ±5.8*	28.7 ±5.9*
	SYTO 9^{+}	15.1 ±0.7	62.4 ±3.4	65.7 ±4.1	48.6 ±17.7	32.9 ±9.3	29.2 ±7.4
	PI^{+}	5.9 ±0.8	21.1 ±5.3	14.3 ±4.4	29.0 ±13.2	39.1 ±2.9	42.2 ±1.3
HM +2%/RM +2%	SYTO 9/PI^{dim}	77.5 ±0.4	22.3 ±9.0	21.0 ±5.5	18.1 ±8.2	7.1 ±2.4	10.4 ±2.6*
	SYTO 9^{+}	15.6 ±0.7	64.6 ±10.9	71.2 ±4.3	60.8 ±5.3	58.6 ±10.6	48.0 ±5.6
	PI^{+}	5.9 ±0.7	13.1 ±4.1	7.8 ±1.8	21.1 ±5.0	34.3 ±10.2	41.6 ±5.1

*Sub-population primarily composed of SYTO 9/PI$^{dim(\beta)}$ sub-population as described in section 5.3.5.

Table 5.5 Percentage of spores (%) in each sub-population after a moderate (96ºC for 24 min) thermal treatment

NaCl Concentration (Heating/Recovery)	Sub-population	time (h)					
		0	**6**	**12**	**24**	**48**	**72**
HM +0%/RM +0%	SYTO 9/PI^{dim}	31.7 ±3.0	20.2 ±10.6	24.9 ±7.9	12.1 ±3.6	12.0 ±3.2	5.9 ±1.0
	SYTO 9^{+}	49.2 ±7.5	60.0 ±7.6	62.6 ±3.9	38.6 ±1.6	8.1 ±0.5	4.1 ±1.2
	PI^{+}	19.1 ±5.2	19.9 ±3.5	12.5 ±3.9	49.4 ±4.7	79.9 ±2.7	89.0 ±2.3
HM +0%/RM +2%	SYTO 9/PI^{dim}	31.6 ±3.2	20.3 ±11.3	27.8 ±8.7	7.9 ±2.1	8.6 ±3.3	5.5 ±1.0
	SYTO 9^{+}	48.8 ±8.9	65.1 ±8.7	64.0 ±5.2	62.2 ±4.8	35.1 ±14.5	15.4 ±9.1
	PI^{+}	19.6 ±5.3	14.6 ±3.1	8.2 ±3.2	29.9 ±5.8	55.4 ±17.7	78.0 ±9.9
HM +2%/RM +0%	SYTO 9/PI^{dim}	38.7 ±1.3	21.0 ±5.4	17.3 ±7.7	15.7 ±1.5*	20.5 ±5.5*	24.5 ±5.3*
	SYTO 9^{+}	40.2 ±5.9	34.0 ±1.5	55.8 ±3.7	39.4 ±3.0	15.7 ±1.9	14.9 ±2.5
	PI^{+}	21.2 ±7.1	45.0 ±3.9	25.9 ±10.1	44.9 ±4.3	62.8 ±8.3	60.6 ±8.7
HM +2%/RM +2%	SYTO 9/PI^{dim}	38.6 ±1.1	21.2 ±5.8	24.4 ±9.4	9.9 ±3.7	9.5 ±1.4	7.6 ±1.7
	SYTO 9^{+}	40.5 ±5.6	54.7 ±9.4	63.5 ±5.1	55.1 ±9.3	45.7 ±3.8	35.0 ±1.6
	PI^{+}	20.9 ±5.7	24.1 ±13.8	12.2 ±3.3	35.0 ±13.0	43.8 ±4.8	55.4 ±0.7

*Sub-population primarily composed of SYTO 9/PI$^{dim(\beta)}$ sub-population as described in section 5.3.5.

Table 5.6 Percentage of spores (%) in each sub-population after a severe (96ºC for 54 min) thermal treatment

NaCl Concentration (Heating/Recovery)	Sub-population	time (h)					
		0	6	12	24	48	72
HM +0%/RM +0%	SYTO 9/PI^{dim}	13.4 ±1.6	5.8 ±0.9	14.7 ±7.7	5.7 ±0.8	5.3 ±0.9	9.3 ±2.6
	SYTO 9^{+}	70.8 ±5.0	64.5 ±10.0	63.1 ±2.7	59.4 ±10.9	51.2 ±15.3	33.2 ±14.6
	PI^{+}	15.9 ±3.4	28.7 ±11.0	22.1 ±9.0	34.0 ±11.0	43.6 ±15.7	57.5 ±15.8
HM +0%/RM +2%	SYTO 9/PI^{dim}	13.2 ±1.5	7.5 ±4.3	12.6 ±5.7	4.3 ±1.2	3.6 ±0.5	4.6 ±0.6
	SYTO 9^{+}	71.1 ±5.2	67.3 ±5.3	70.9 ±2.3	73.5 ±5.7	64.4 ±13.0	47.0 ±15.1
	PI^{+}	15.7 ±3.7	25.2 ±10.3	15.5 ±5.0	22.2 ±5.6	32.1 ±12.7	48.4 ±15.7
HM +2%/RM +0%	SYTO 9/PI^{dim}	9.0 ±1.7	10.8 ±2.9	15.9 ±12.1	3.2 ±0.5	8.2 ±3.1	9.6 ±2.5
	SYTO 9^{+}	45.7 ±15.9	45.9 ±12.6	42.4 ±8.8	42.4 ±15.2	42.8 ±13.0	35.8 ±14.8
	PI^{+}	44.3 ±17.5	43.3 ±15.5	41.7 ±19.6	54.4 ±15.2	49.0 ±15.1	53.6 ±17.0
HM +2%/RM +2%	SYTO 9/PI^{dim}	14.3 ±5.2	8.2 ±2.9	11.3 ±7.0	7.5 ±3.2	5.8 ±2.1	7.8 ±3.3
	SYTO 9^{+}	42.5 ±13.9	51.6 ±15.7	54.0 ±12.0	51.4 ±18.4	45.7 ±20.4	45.2 ±13.9
	PI^{+}	43.2 ±20.0	40.2 ±18.7	34.6 ±17.2	41.0 ±21.4	47.5 ±22.4	45.9 ±17.0

5.4 Discussion

This study provided a number of findings contributing to a mechanistic understanding of *C. sporogenes* spore thermal inactivation. Firstly, it validated the suitability of the fluorophores SYTO 9 and PI in combination with FCM to both quantify and discriminate between individual *C. sporogenes* spores of varying physiological states. Secondly, by applying the aforementioned methodology the effect of varying NaCl concentration during a heat treatment on both the likelihood of spore thermal injury and also the type of structural injury sustained was described. Finally, FCM was applied to investigate post-heating process changes in spores to estimate the likelihood of germination after the process and investigate the sublethal injury of thermally-injured spores. A simplified model describing these effects observed in this study is shown in Figure 5.5.

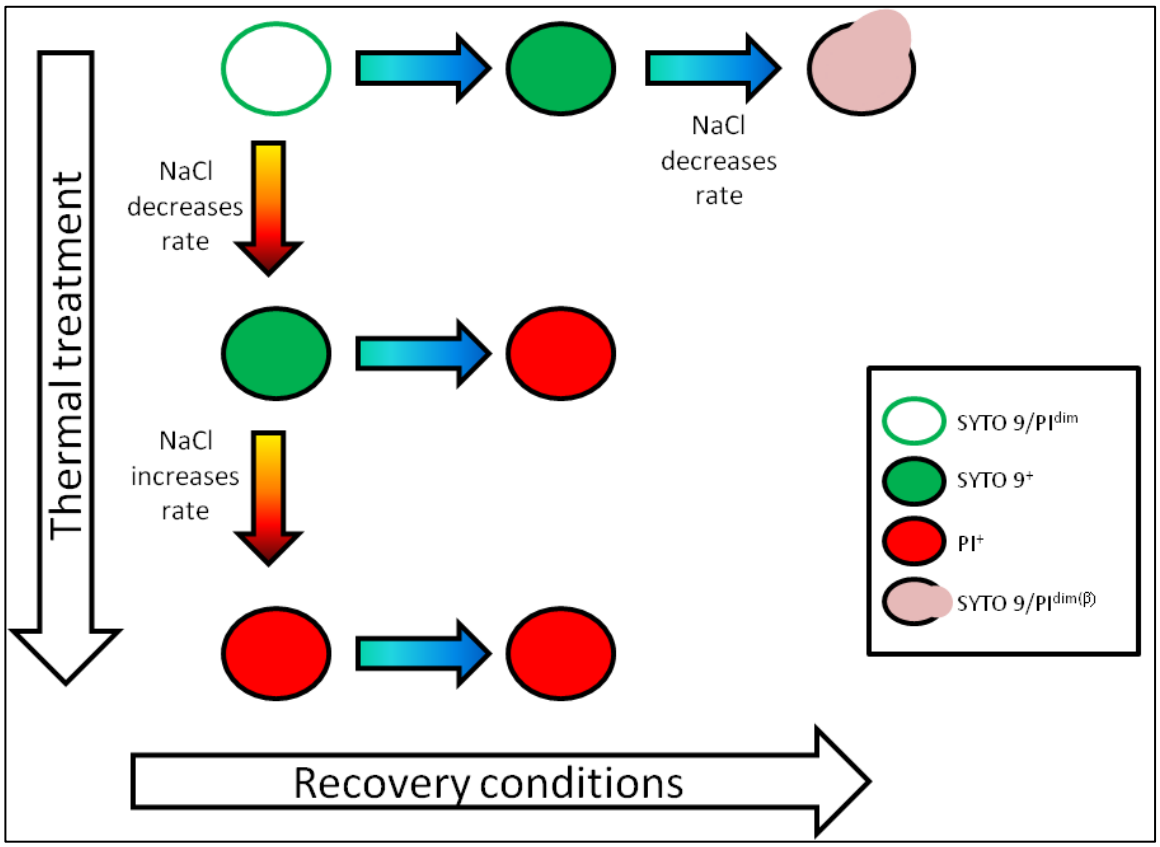

Figure 5.5 Simplified model indicating the likely spore transitions and effect of increasing NaCl concentrations on these transitions observed in this study. SYTO 9/PI^{dim} transition to a SYTO 9^{+} and then PI^{+} state occurs as the thermal treatment extends. Dependent on the length of the recovery period and initial state of the spores at the beginning of the recovery period spores may transition through multiple states to a resting SYTO 9/$PI^{dim(\beta)}$ state.

5.4.1 Permeability to the fluorescence stains SYTO 9 and PI can be used to discriminate between spores with varying physiologic states.

FCM effectively discriminated between spores exhibiting differing structural states. Based on a combination of the permeability of the dyes and the fluorescence properties of control populations spores were grouped based on fluorescence emissions and structural states predicted (Table 5.2).

Dormant spores exhibited relatively low levels of emission of both SYTO 9 and PI and were characterised as SYTO 9/PI^{dim} spores. The weak fluorescence was indicative that neither stain could reach the spore core to intercalate with nucleic acid. In addition the higher light scattering of dormant spores was indicative of a highly refractile body and a dehydrated core in these spores. Lack of permeability to stains and larger chemical molecules is considered one of the characteristics of dormant spores (Setlow, 2003). Amongst a number of proposed theories in regards to the impermeability of dyes to the core of dormant spores are low inner membrane permeability (Cowan et al., 2004), saturation of α/β small acid soluble proteins (SASP) preventing nucleic acid stains from binding with target DNA (Kong et al., 2010) and the hydrophobic nature of the spore core (Setlow, 2003).

Despite the apparent exclusion of the fluorogenic dyes from the spore core a relatively weak fluorescence emission of SYTO 9 was observed in the SYTO 9/PI^{dim} spores. This weak SYTO 9 emission was indicative of non-specific staining of SYTO 9 external to the spore core, previously observed as a concentric ring in *Bacillus* spores (Melly et al., 2002; Setlow et al., 2002). The exact location of the external staining of SYTO 9 relevant to the spores is unknown; however it has been proposed that SYTO 9 may be constrained within hydrophobic domains in the protein coat of *Bacillus* spores (Ferencko and Rotman, 2010).

The increased green fluorescence in germinating spores compared to dormant spores was indicative that the germinating spores were able to uptake SYTO 9 to the spore core allowing it to bind to nucleic acid. Kong et al. (2010) have illustrated that germinating spores permit the ingress of the similar acting green fluorescent nucleic acid stain SYTO 16 and that the majority of the fluorescence associated with SYTO 16 uptake can be attributed to binding of the dye to DNA in the spore core. A similar

response has previously been shown by the uptake of ethidium bromide (Coote et al., 1995 and Anderson, 1997), DAPI (Ragkousi et al., 2000) and also non-fluorescent stains (Hamouda et al., 2002; Setlow, 2003). These germinating spores exhibited similar light scatter characteristics, despite having started germination they still retained refractile characteristics. This was indicative of only partial rehydration of the core and suggested that these spores have not completed all the degradative steps associated with fully germinated spores.

A heat-inactivated control spore preparation produced the highest SYTO 9 fluorescence observed in the control heat-inactivated population. An upshift in red fluorescence occurred concomitantly with the increase in green fluorescence, and it is probable that the upshift in red fluorescence observed is attributable to fluorescence resonance energy transfer of SYTO 9 rather than PI staining (Stocks, 2004). The increased SYTO 9 fluorescence observed in heat inactivated spores could be attributed to denaturation of nucleic acid binding proteins, as spores deficient in DNA binding SASP had increased levels of bound nucleic acid dyes (Kong et al., 2010). Another possibility is that the heat inactivation process applied resulted in degradation of the spore cortex resulting in increased penetration of SYTO 9. The higher fluorescence emission of SYTO 9 in wet heat inactivated spores compared to germinating spores is in contrast to the results of Zhang et al., (2011) who in spores of multiple *Bacillus* spp. observed a 15-fold decrease in the fluorescence emission of the similar acting SYTO 16 nucleic acid stain in wet heat inactivated spores compared to germinating spores. This variation can be attributed to differences in the analytical approaches, species variation and/or differences in germination stages between spores.

5.4.2 Flow cytometric derived description of the thermal inactivation of spores of *C. sporogenes*

Heat-induced formation of SYTO 9^+ and PI^+ populations during the thermal process was indicative that a population of spores does not simply exist as dormant or heat inactivated. This heterogeneous response of spores can be explained by a number of hypotheses: i) that the heat inactivation of spores is a two-stage process of initial permeability to SYTO 9 followed by disruption of the membrane; ii) that spores are

simultaneously heat activated to begin germination and also heat inactivated as two independent first-order reactions; or iii) that within the initial homogeneous population exist heat-sensitive sub-populations. It is likely that the observed heterogeneity is highly complex and all three phenomena occur to varying degrees. The results of this study are most supportive of a model that the integrity of dormant spores prevents the ingress of SYTO 9 prior to irreversible disruption of the membrane (see Figure 5.5 for a simplified model).

Based on this description, the most effective method to describe the stochastic response of individual spores to a thermal process is to monitor the reduction of SYTO 9/PI^{dim} spores. However, it is important to note that these spores were not all capable of outgrowth on CSRA. This was exemplified in the moderate treatment which had a ~99.99% reduction in recoverable spores compared to only a ~70% reduction in SYTO 9/PI^{dim} spores (Table 5.3).

Another key finding is that thermally inactivated spores in the unadulterated heating medium displayed only low levels of PI^{+} spores (Figure 5.4). This was indicative that the thermal treatment applied was not sufficient enough to compromise the membrane integrity of the spores to allow the ingress of enough PI to completely dissociate SYTO 9 from the nucleic acid. Consequently, it can be concluded that loss of culturability was not contingent on the occurrence of complete loss of membrane integrity. Coleman et al. (2007) have suggested that membrane changes resulting in permeability to SYTO dyes and PI occur well after culturability is lost.

5.4.3 Protective effect of NaCl on spores during a thermal process is time dependent

The addition of NaCl to the heating menstruum reduced the percentage of spores that were permeable to SYTO 9 following a short thermal treatment, suggesting that NaCl had a protective effect by limiting the percentage of spores that became rehydrated. The rehydration of spores has been associated with the loss of spore heat resistance (Setlow, 2003). It can then be hypothesised that NaCl-induced inhibition of rehydration may be a factor resulting in increased survival of spores at short treatment times. NaCl-induced inhibition of the rehydration of spores has been reported at NaCl

concentrations as low as 1% (Duncan and Foster, 1968, Montville, 1984). Conversely, higher concentrations of NaCl in the heating menstruum were associated with higher numbers of spores in the PI^+ population suffering a form of temperature-induced membrane damage when exposed to the most severe thermal treatment.

Coupled together these results suggest that whilst NaCl may initially contribute to the protection of spores against thermal stress it may represent an additional stress to spores heated over longer durations (Figure 5.5). Thus the thermal inactivation of spores in the presence of hurdles such as NaCl is unlikely to be log-linear and time-independent. This has important implications in the spore prediction using time-independent models (i.e. D-values) which may lead to an inadvertent underestimation or overestimation of the effect of the combined hurdles. Thermal inactivation models describing the dependence of the rate or probability of spore inactivation on the state of the spore (Shull et al., 1963; Rodriguez et al., 1992; Sapru et al. 1993; Corradini et al., 2010) are more likely to be appropriate for describing the thermal inactivation of spores in the presence of additional hurdles such as NaCl which may alter the rate of spore germination.

5.4.4 Spore structural changes post heat treatment under recovery conditions (+/- NaCl)

Heat-treated spores placed in a recovery medium indicated significant variation based on the length of heat treatment applied and the response of individual spores appeared highly dependent on the physiological state after the treatment (Figure 5.5). Spores that were either in SYTO 9+ or PI+ state were highly unlikely to transition further beyond a resting state in one of these populations as indicated by the severely heat-treated spore populations (Tables 5.4-5.6). Spore populations which received milder thermal treatments exhibited a much more complicated transition typically from SYTO $9/PI^{dim} \rightarrow$SYTO $9^+ \rightarrow PI^+ \rightarrow$ SYTO $9/PI^{dim(\beta)}$. These results in general are consistent with multistage degradative steps associated with spore outgrowth (Setlow 2003). Spores undergo at least two stages of germination before undergoing outgrowth, the first stage involves significant excretion of ions and dipicolonic acid and partial hydration of the spore core, followed by the second stage which involves

degradation of the spore cortex, and full hydration of the spore core prior to outgrowth (Figure 2.3).

The SYTO 9/PI$^{dim(\beta)}$ state was observed within 72 h after a mild or moderate thermal treatment in recovery medium with no added NaCl. The identity of this population which exhibited similar fluorescence emissions to dormant spores while exhibiting lower light scatter characteristics still remain somewhat speculative. The presence of this population under only maximal survival conditions is indicative that it is most likely related to an outgrowth related phenotype which has been prevented from outgrowing due to the addition of ampicillin to the outgrowth medium. The weak fluorescence attributable to this population could be ascribed to a number of possibilities including structural modifications in the binding potential of the DNA or RNA in these spores. These changes may interfere with the binding potential of SYTO 9 to the DNA and hence decrease the fluorescence of these spores. Alternatively, it may be due to non-specific binding of SYTO 9 to proteins abundant in the spore core (Setlow, 2007).

In all cases, the addition of 2% NaCl to the recovery medium increased the percentage of spores in a rehydrated state (SYTO 9+), indicating a decrease in their ability to transition beyond germination. The addition of NaCl to recovery media is widely used as an indicator of sublethal injury, it has repeatedly been shown to decrease the number of spores recovered on plated media from heat treated populations (Roberts et al., 1966; Montville, 1984; Hutton et al., 1991; Coleman and Setlow, 2009). Our results suggest that NaCl acts upon heat treated cells not by decreasing the rate of germination (as indicated by the decrease in SYTO 9/PI^{dim} spores) but by preventing the transition beyond the secondary rehydrated state (SYTO 9^{+}). This is consistent with results of Webb et al., (2007) who observed increased variability in germination and outgrowth times of Group II *C. botulinum* spores in the presence of an additional 2% NaCl. Coleman and Setlow (2009) have also suggested that decreases in the viability of heat treated *Bacillus* spores at similar NaCl concentrations as employed in this study could be ascribed to a defect in metabolic capability rather than an inability to germinate or DNA damage. This is supported by the results of this study which suggest a rate limiting effect of NaCl for post germination processes rather than a role in the prevention of rehydration.

5.5 Concluding remarks

The results of this study confirmed ability of FCM with SYTO and PI to differentiate between spores in varying physiological states. From this it is possible to describe and quantify the heterogeneity present during thermal inactivation of spores in the presence of NaCl, a common stressor encountered in food preservation systems. FCM observations of spores in response to a thermal treatment allowed the differentiation between spores in various physiological states at the individual cell level, quantified the proportion of the spores in these states at any one time and separated the independent effects of increased NaCl during and post heating.

During the heating, increasing the NaCl concentration up to 4% initially protected spores by preventing the transition of spores into a rehydrated state, however it increased the level of membrane associated damage to those spores that did rehydrate. Over a 72 h recovery period after the thermal treatments, increased NaCl decreased the proportion of spores transitioning beyond germination. Further studies are required to determine whether this equates with a prevention of outgrowth. These results indicate that the effect of inhibitory solutes such as NaCl on the heterogeneity of spore states should be considered in designing both process and product criteria to meet food safety objectives. Further to this, a greater understanding and appreciation of heterogeneity of spore states will improve predictions and models of spore thermal inactivation in food preservation systems.

Further characterisation of individual spores is essential in the development of more accurate predictive models of spore inactivation. The results of this study were based purely on discriminating spore based structural variation allowing changes in permeability to SYTO 9 and PI. Within these populations it is likely that even further diversity exists. For example, germinating spores and moderately heat-inactivated spores may both exhibit similar levels of permeability to SYTO 9 whilst still excluding PI. The use of alternative fluorescent probes targeting different physiological characteristics is one approach that may provide further insights into the heterogeneity of these populations.

Chapter 6 – Ultrastructure variation in heat sensitive and heat resistant spores of *Clostridium sporogenes*

6.1 Introduction

The reported greater heat resistance of *C. sporogenes* str. PA 3679 spores compared to *C. botulinum* is a distinguishing characteristic of the strain. Previous chapters introduced prospective genetic contributors to heat resistance (Chapter 3) and also described the heterogeneous response of individual spores to a thermal process (Chapter 5). This chapter describes differences observed in the heat resistance and ultrastructure of spores produced from two strains of *C. sporogenes* PA 3679 received from different culture collections (NCTC 8594 and ATCC 7795).

Extreme tolerance to heat is one of the defining features of bacterial spores. Considerable effort is taken in defining the heat resistance of spores formed by bacteria associated with foodborne spoilage and disease. This parameter, described through D-values, is often employed in the derivation of risk-based assessments in regards to the stability and safety of thermally processed food commodities. As such, any variability between heat resistance values may significantly influence thermal processing schedules and/or performance objectives.

A significant challenge for research laboratories is the batch variation often observed in spore crops (Mah et al., 2008a). For example, the heat resistance values of *C. sporogenes* spores exhibit significant variability, with reported D_{121} values ranging from 0.1-1.5 min (Mah et al., 2008a; Brown et al., 2012; Diao et al, 2014). This is undoubtedly in part due to the multitude of factors that have been shown to influence the heat resistance of spores, including but not limited to species specific variation, sporulation temperature, sporulation medium composition (Waites et al., 1979), and spore storage conditions (Mah et al., 2008b). Such factors all influence in particular components of the physical architecture such as the spore core, cortex and coat.

A correlation between heat resistance and variation in core and cortex size has been well established in the spores. The ratio of spore core to sporoplast (core and cortex) volume (Beaman et al., 1982) and the core wet density (Nakashio and Gerhardt, 1985) have both been associated with spore heat resistance in a variety of *Bacillus* spp. Central to spore heat resistance is a dehydrated spore core. The mechanism of maintaining a dehydrated spore core is proposed to be associated with core contraction and expansion of the cortex (Gould and Dring, 1975). Fewer studies have been conducted with clostridial spores, however increased coat layer and cortex size have been correlated with increased heat resistance in *C. sporogenes* PA 3679 spore crops (Mah et al., 2008a).

Given the significance of the spore structure, the objective of this study was to characterise variation in the spore structure and genomes of two *C. sporogenes* strains exhibiting differential heat resistance. In this chapter, structural variation of spores is described using transmission electron microscopy (TEM) and flow cytometry. Furthermore, variation in the genomes between strains is described. Together such studies may further elucidate factors contributing to variation in the heat resistance of clostridial spores.

6.2 Materials and Methods

6.2.1 Spore crop preparations of *C. sporogenes* PA 3679 spores

Cultures of *C. sporogenes* NCTC 8594 and ATCC 7955, both described as PA 3679, were received from the National Collection of Type Cultures and American Type Culture Collection respectively. Spores crops were grown and purified as previously described (Chapter 5, sections 5.2.1-5.2.3). Three batches of spore crops were prepared from the same initial glycerol stock for each culture.

Spores were purified in a 70% Urografin® solution prior to suspension in SDW to a concentration of *ca.* 10^7 spores mL^{-1} and stored at -80°C prior to use.

6.2.3 Heat resistance measurements and calculations

Inactivation curves of spore crops were determined at 90.0±0.2°C, 96±0.2°C and 105±0.2°C in 0.2M potassium phosphate buffer (pH=6.9). Temperatures were monitored by a thermocouple placed in control tubes and were recorded on a 1050 series Squirrel Data Logger (Grant, Cambridgeshire, UK). The initial spore concentration was determined after heat activation for 10 min at 70°C.

Heating experiments at 90°C and 96°C were conducted in glass test tubes in a circulating water bath (Thermoline, Wetherill Park, Australia). Test tubes containing 9.9 mL of heating menstrua were immersed into the water bath ensuring that the level of heating menstruum was below the water line. After equilibration to the appropriate temperature, a 100 μL aliquot of defrosted spore preparation was pipetted into the centre of the tube using an 89mm pipette tip. After the required heating time elapsed, spore preparations were rapidly chilled in an ice water mixture.

Heating experiments at 105°C were conducted using the method of Kooiman and Geers (1975), with some modifications. Custom made 10mL stainless steel tubes, fitted with 12.5 mm rubber septa in the cap (Alltech, Dandenong, Australia) were filled with 9.9 mL of heating menstrua. These tubes were immersed into an oil bath (< 5mm below the surface) containing M bath fluid (Julabo, Seelbach, Germany). After equilibration 100 μL of defrosted spores was injected into the centre of the tube using

a 22G, 3 inch spinal needle (BD Biosciences). After the required heating time elapsed, spore preparations were rapidly chilled in an ice water mixture prior to enumeration.

Trial experiments indicated that the injection of 100 μL inoculum decreased the heating menstruum temperature by no more than 0.5°C and the come-up time to return to a temperature ±0.2°C was within 2 minutes. The cool-down rate ranged between 8.0°C and 10.0°C s^{-1} in the ice slurry (data not shown).

Spores were enumerated by pour plates on a modified version of PA 3679 agar (MPAA), (Grischy et al., 1983). The medium consisted of 10 g L^{-1} of Tryptone (Oxoid), 2 g L^{-1} soluble starch (Ajax), 2 g L^{-1} Yeast Extract (Oxoid) and 15 g L^{-1} no. 1 Bacteriological agar (Oxoid). Immediately prior to pouring plates the agar was supplemented with a filter sterilised solution (0.2 μM filter, Millipore) to a final concentration of 1.32 g L^{-1} sodium thioglycollate (Ajax), 1 g L^{-1} $NaHCO_3$ (Ajax) and 0.4mg L^{-1} Lysozyme (Sigma Aldrich).

The survivor curves were plotted as $\log_{10}$ cfu mL^{-1} versus time. The D_T was calculated from the negative reciprocal of the slopes of the regression lines, using the linear portion of the survivor curves. This resulted in the exclusion of the last datapoint point for the ATCC strain at 90°C, both strains at 96 °C and the NCTC strain at 105°C.

6.2.4 Spore staining and flow cytometry

Spores that were subjected to 96±0.2°C thermal treatment were analysed using flow cytometry. Preliminary trials indicated that spores treated at 90°C and 105°C were poorly resolved and it was not possible to discriminate clear subpopulations. Staining of spores was as previously described in section 5.2.7. The flow cytometer settings were as described in section 5.2.8. A minimum of 10,000 individual spores were acquired for downstream analysis for each experiment.

Analysis was conducted as described in section 5.2.9, with some modifications of the gating strategy. Initially, non-spore-related events were removed by gating on the basis of FSC-H and FL1-H, followed by manual selection of “dye permeable” and “dye impermeable” populations on the basis of FL1-H and FL3-H.

6.2.5 TEM sample preparation

Spore suspension were suspended in 1mL of 2% paraformaldehyde (Sigma Aldrich) and 2% glutaraldehyde (Sigma Aldrich) prepared in 0.1M phosphate buffer (pH 7.2) for 2 hours at 4°C. Spores were washed three times for 10 minutes with buffer prior to embedding in 1% low melt agarose. The pellets were then post fixed with 1% osmimum tetroxides (Ajax) in buffer for 1 h at room temperature, followed by 3 x 10 minute washes with the buffer. The pellet was then dehydrated in graded ethanol series of 20%, 30% (stained *en bloc* with 2% uranyl acetate during this step), 50%, 70%, 90%, 95% and 100% ethanol (three times) for 20 minutes per step.

After dehydration spores were infiltrated with LR white resin (Sigma Aldrich) by incremental 3:1, 2:1, 1:1, 1:2, 1:3 pure ethanol:resin mixtures for 1 h per step, prior to suspension into pure resin overnight under vacuum. Samples were then resuspended in fresh pure resin for 1 h prior to polymerisation overnight at 60°C. Thin sections were prepared using a Reichert ultracut ultramicrotome (Reichert, NY, USA), and collected on copper grids (200 mesh) (ProSciTech, Qld, Australia).

Semi-thin sections (1μm) were initially visualised by staining with methylene blue. Ultrathin sections (70nm) were stained with aqueous uranyl acetate for 30 minutes then in lead citrate for 4 minutes and viewed on a Philips LM 10 transmission electron microscope (Philips, NY, USA).

6.2.6 Measurement of spore size

Measurements of spores were taken using ImageJ software v1.47 (http://rsbweb.nih.gov/ij.download.html). Measurements were made on all spores for the exosporium, outer coat, outer cortex and core. The total area of each structure was calculated using the "Analyze particle" function in ImageJ. A minimum 100 spores were measured.

Due to the nature of TEM sectioning through different positions and angles of spores, absolute measurements were not compared. Instead the ratio of areas to the sporoplast was used. Any spores that did not meet a particle size criterion (>10000 pixels) and upon manual inspection did not display clear differentiation between the core and cortex were disregarded from analysis.

6.2.7 Comparative genomics of strains

MiSeq reads of *C. sporogenes* ATCC 7955 (described in Chapter 4) were aligned against the pseudomolecule of *C. sporogenes* NCTC 8594 (constructed in Chapter 3). Reads were mapped against a reference genome using BWA-MEM v0.7.4 (Li, 2013). In order to reduce downstream computational requirements the resultant SAM file was compressed into a BAM file using SAMtools v0.1.19 (http://samtools.sourceforge.net). The final quality of the alignment was manually assessed by visualisation in Geneious (v6.1.4; http://www.geneious.com/).

6.3 Results

6.3.1 Heat inactivation profiles

Heat inactivation curves at 90°C, 96°C and 105°C were prepared for spore crops of *C sporogenes* NCTC 8594 and *C. sporogenes* ATCC 7955 (Figure 6.1). All curves displayed tailing as the number of survivors decreased towards 10 cfu mL^{-1}. Consequently, the D_T-values for each of these temperatures were calculated excluding the tails of the curve.

The spores of *C. sporogenes* ATCC 7955 had higher or similar heat resistance compared to those of *C. sporogenes* NCTC 8594 under all conditions (Table 6.1). At the lowest challenge temperature no significant difference in D_{90} values between the ATCC spores and NCTC spores was observed. At the higher two temperatures the D_{96} and D_{105} values of the ATCC strain were 11.7 and 2.9 min compared to the respective values of 6.5 and 1.3 min for the NCTC spores. The difference between heat resistance profiles increased with increasing temperatures. Plotting the D_T-values against the temperatures (Figure 6.2) the z-values of the ATCC and NCTC spore crops were calculated as 15.6 and 11.4, respectively (Table 6.1).

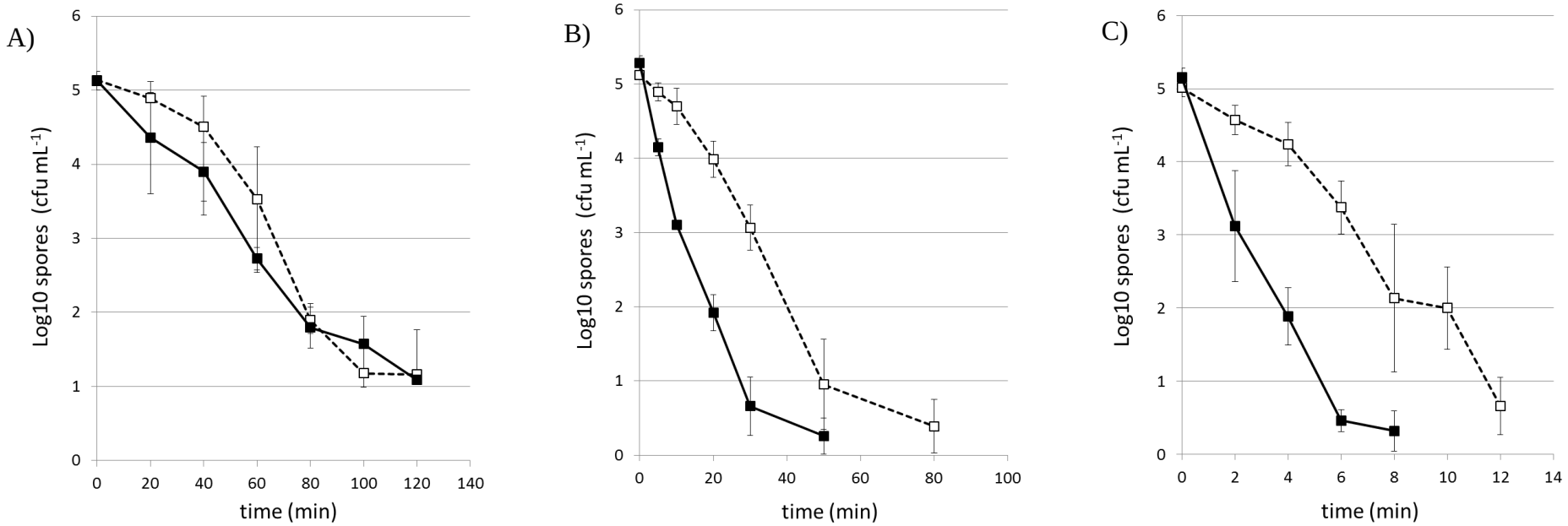

Figure 6.1 Heat inactivation of *C. sporogenes* ATCC 7955 (□) and NCTC 8594(■) spores at 90°C (A), 96°C (B) and 105°C (C) in 0.2M potassium phosphate buffer (pH 6.9).

Table 6.1 Summary of heat resistance properties of *C. sporogenes* ATCC 7955 and NCTC 8594 spore crops.

	ATCC	NCTC
HEAT RESISTANCE		
- D_{90}	23.8±4.5	25.9±2.1
- D_{96}	11.7±1.5[a]	6.5±0.4[a]
- D_{105}	2.9±0.7[a]	1.3±0.1[a]
- z-value	16.3	11.5
-estimated $D_{121.1}$	0.31	0.05
- Maximum rate of spore permeability to fluorophores at 96ºC (%/min)	6.33±0.88[a]	1.63±0.49[a]

[a] indicates significant difference between samples ($p>0.05$)

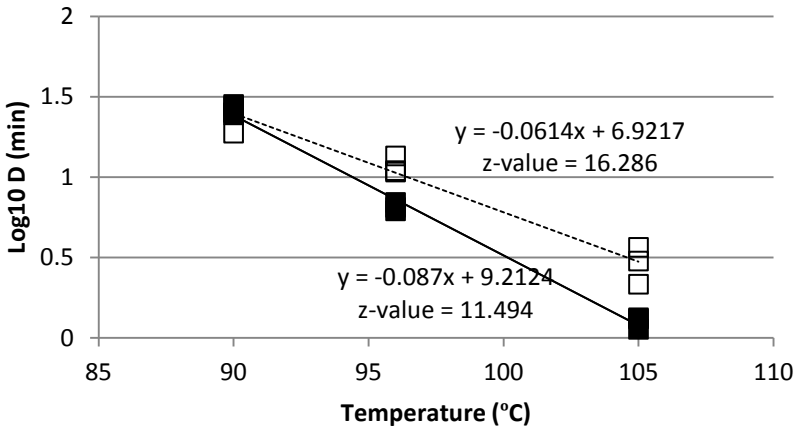

Figure 6.2 Derivation of z-value of *C. sporogenes* strains (■, NCTC 8594; □, ATCC 7955)

6.3.2 Flow cytometric evaluation of spores

Spores that were heat treated at 96°C were assessed for susceptibility to temperature induced structural changes, permitting staining by the flurophores SYTO 9 and PI (Figure 6.3). Spores were assessed as "dye impermeable" or "dye permeable" dependent on their fluorescence characteristics. Dormant spores were shown to

exhibit low background levels of fluorescence and congregate in the “dye impermeable” region of the flow cytometric dotplots (Figure 6.3, green boxes). As thermal treatments extended the percentage of “dye permeable” spores (Figure 6.3, red boxes) was observed to increase.

A more significant shoulder was observed in flow cytometric compared to the inactivation curves prepared from plated recovery (Figures 6.1 and 6.4). Quantification of spores revealed that the ATCC strain became permeable to dyes at a significantly reduced maximum rate (1.63% per minute) compared to the NCTC strain (6.33% per minute) (Table 6.1). This most significant difference between spore populations was observed after a 10 minute period with the NCTC spores having 88% of spores exhibiting permeability to the fluorescent dyes compared to 12% of ATCC spores. This difference decreased with increasing duration until both spore crops exhibited >95% spores permeable to the fluorescence dyes after 80 minutes.

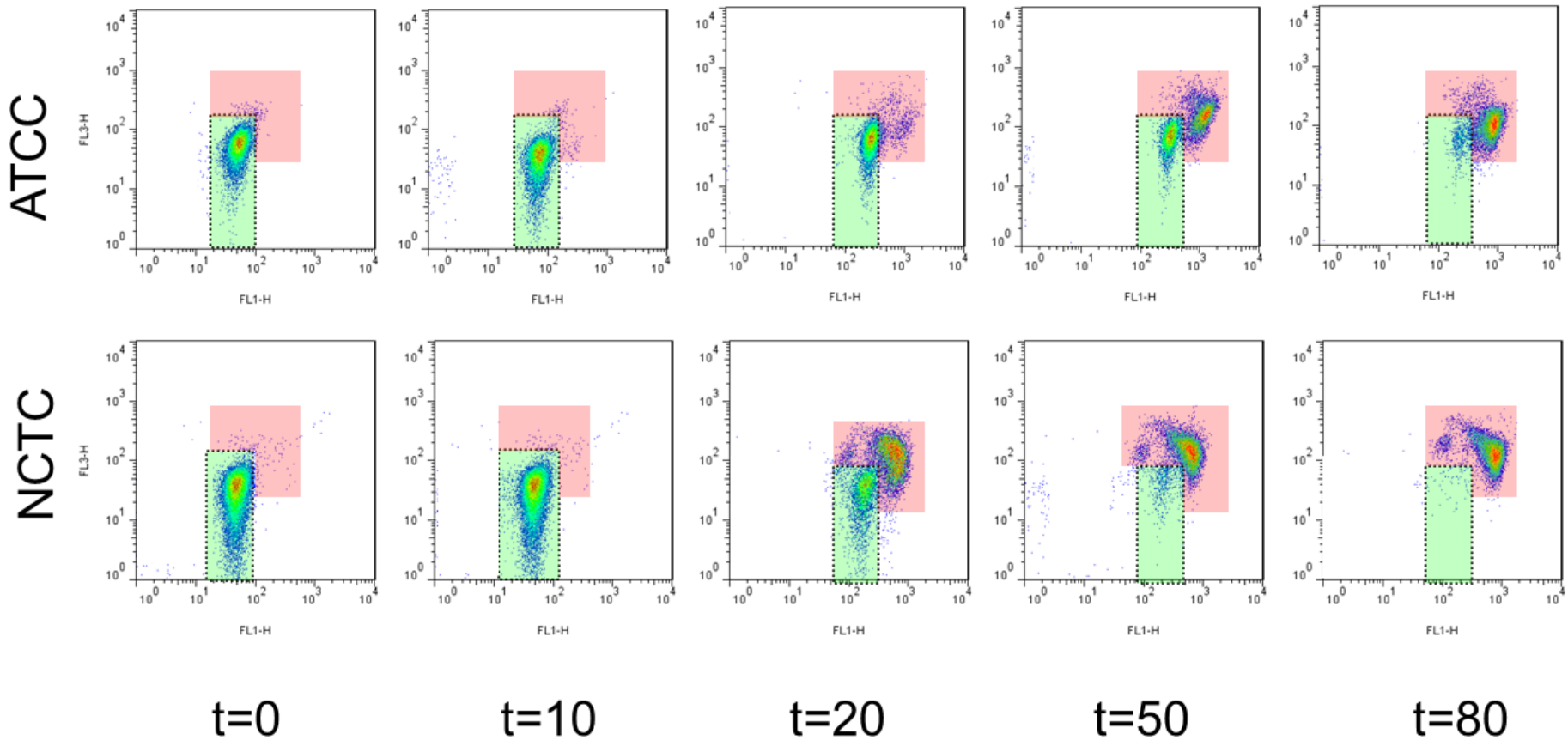

Figure 6.3 Example FCM derived dotplots of spores after exposure to a heat treatment at 96°C for t=0, 10, 20, 50 and 80 minutes. The population present in the green box represent spores that did not permit the ingress of the fluorophores SYTO 9 and PI. The population present in the red box represents spores that permitted the ingress of at least one of these dyes.

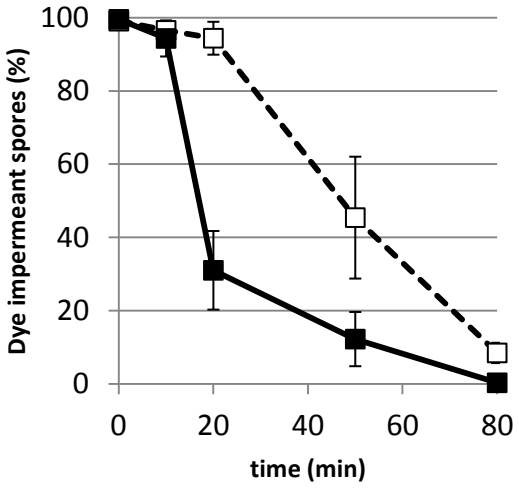

Figure 6.4 Changes in the percentage of spores impermeable to the fluorophores SYTO 9 and PI as measured by flow cytometry. (■, NCTC 8594; □, ATCC 7955)

6.3.3 TEM observations of dormant, activated and heat-treated spores

Spores in a dormant and heat treated state were assessed for differences in the spore ultrastructures via TEM. All spores revealed longitudinal profiles with familiar concentric ring structures, and surrounded by a loose fitting sac-like exosporium (Figure 6.5). Dormant spores exhibited structures of defined concentric rings expanding from a dark central core. The cortex, electron dense outer coat and exosporium were well defined in all micrographs. Dependent on the sectioning plane, various micrographs also exhibited less well defined inner membranes, inner coats and outer membranes. The exosporium exhibited a clearly defined lamellar structure with no clearly discernable external appendages observed (Figure 6.6).

The most distinguishing difference between the ATCC and NCTC spores was the size of the exosporium. The NCTC spores had a constricted exosporium compared to ATCC spores (Figure 6.7, A-B). The more heat resistant spores also had a significantly broader distribution of exosporium size compared to the low heat resistance crops (Figure 6.8). This observation was confirmed by measurement of the ratio of exosporium area to sporoplast area. The mean exosporium: sporoplast ratio

for the ATCC spore crop was ~3:1 compared to ~2:1 for the NCTC spores (Table 6.2).

Size measurements were compared between dormant spore crops. No significant difference was observed in the diameter of the sporoplast size between spore crops, however a significant difference in the mean size of the core was observed. The core diameter of ATCC spores was significantly lower (419.55nm) compared to the NCTC spores (468.88nm) (Table 6.2).

Few discernible differences were observed between dormant and activated spores regardless of spore crops (Figure 6.7, A-D). However differences in core size and shape were more readily apparent after a thermal treatment (Figure 6.7, E-F). Heat inactivated spores were observed to have an enlarged core and decreased cortex which was more easily permeated by the fixatives.

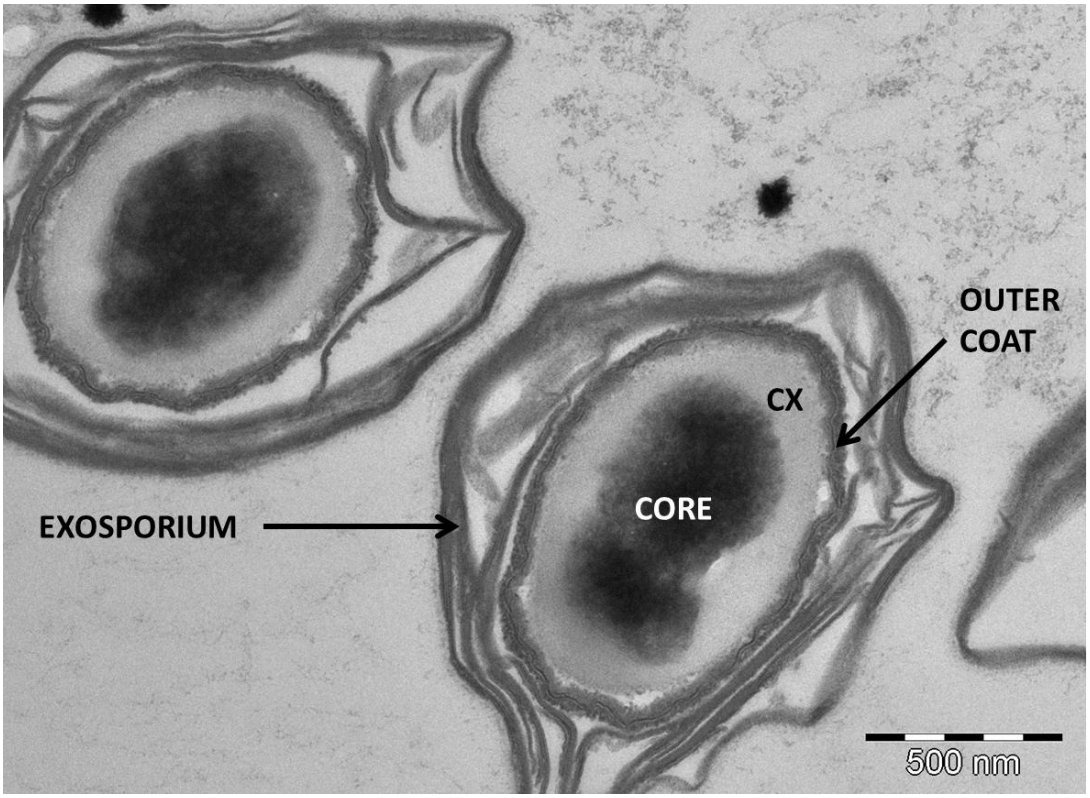

Figure 6.5 Representative TEM micrograph of dormant *C. sporogenes* ATCC 7955 spores. Ultrastructure components measured include the exosporium, outer coat, cortex (CX) and core. Scale bar: 500nm.

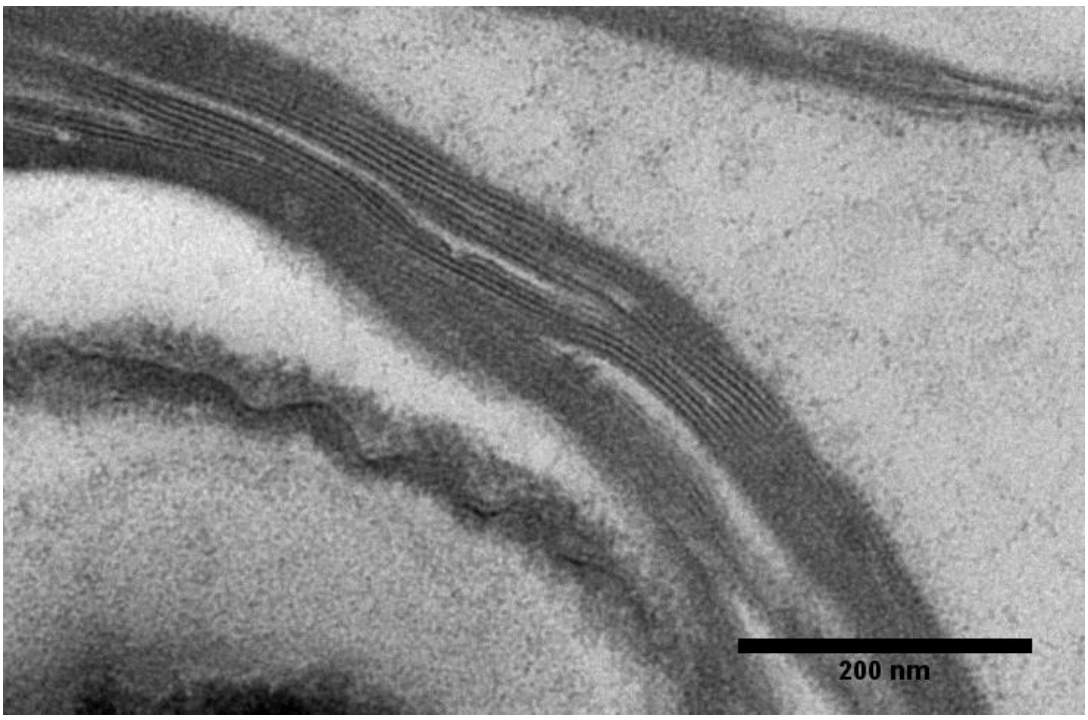

Figure 6.6 Representative TEM micrograph of exosporium observed in *C. sporogenes* ATCC 7955 spores. Scale bar: 200nm.

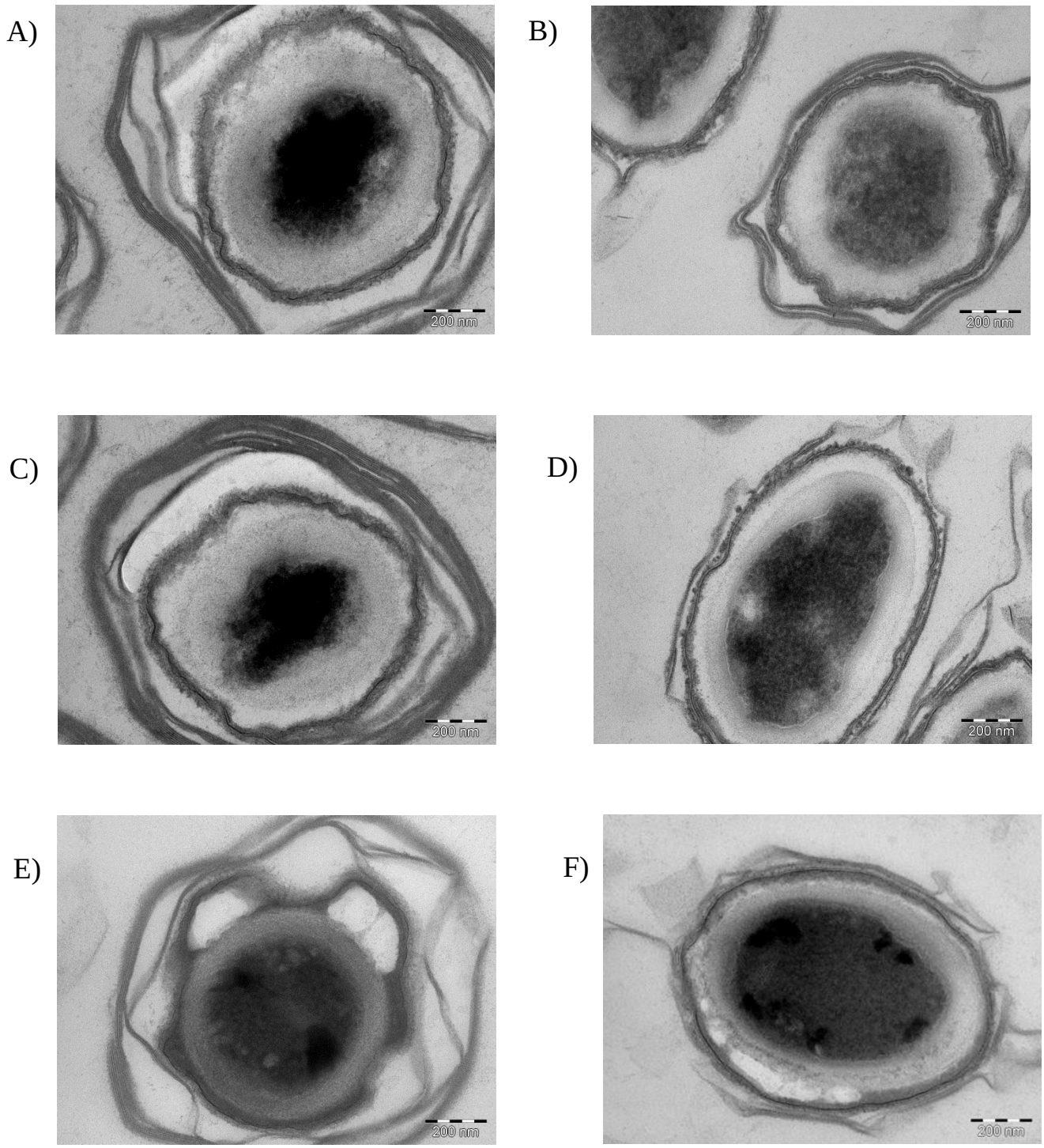

Figure 6.7 Transmission electron micrographs of sectioned spores of *C. sporogenes* ATCC 7955 and NCTC 8594. Treatments: Dormant spores: ATCC 7955 (A) and NCTC 8594 (B); Heat activated spores (70°C for 10 minutes): ATCC 7955 (C) and NCTC 8594 (D); Heat inactivated spores (96°C for 1 h): ATCC 7955 (E) and NCTC 8594 (F). Scale bar: 200nm.

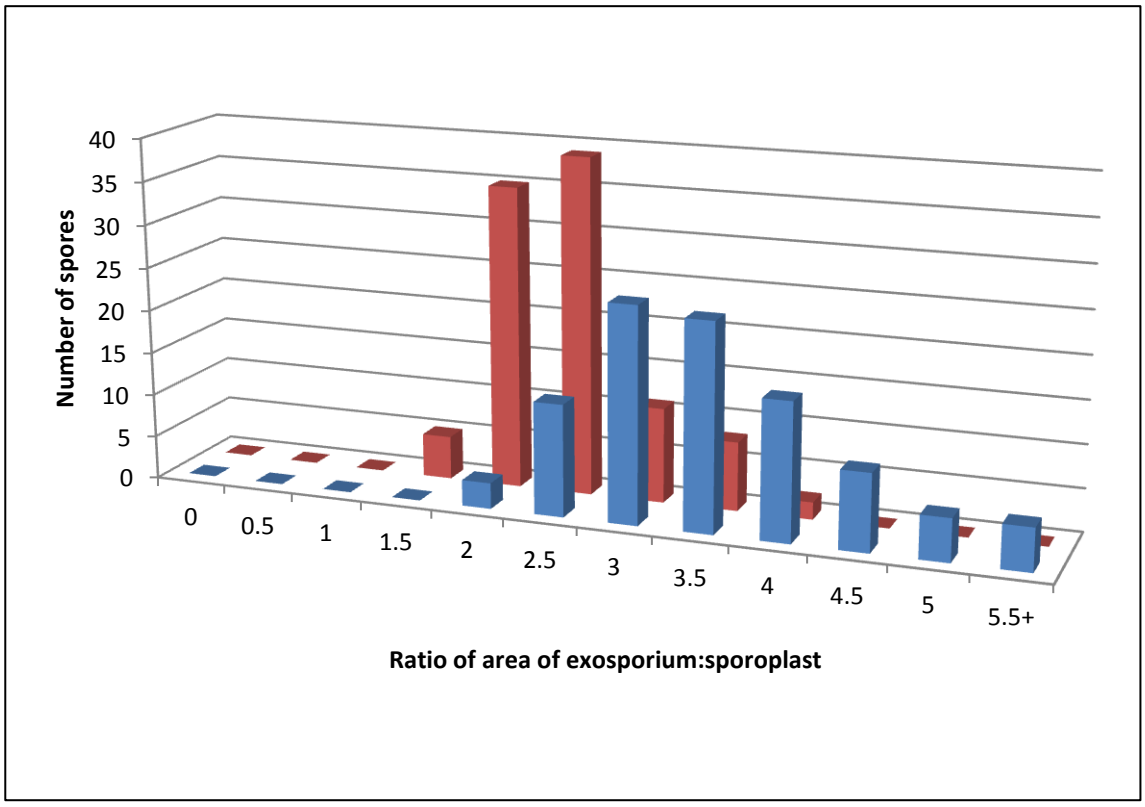

Figure 6.8 Distribution of the exosporium size of *C. sporogenes* ATCC 7955 (blue) and *C. sporogenes* NCTC 8594 (red spores). (n=100).

Table 6.2 Summary of spore structure size of *C. sporogenes* ATCC 7955 and NCTC 8594 dormant spores (n=100) Similar letters indicate no significant difference between samples (p>0.05).

SPORE SIZE		
- Core diameter (nm)	419.55±86.70[a]	468.88±70.39[b]
- Sporoplast diameter (nm)	601.65±90.60[a]	613.05±78.22[a]
- Ratio of the area (nm^2) exosporium: sporoplast	3.30±0.88[a]	2.19±0.49[b]

6.3.4 Reference mapping of *C. sporogenes* ATCC 7955 against the pseudomolecule of NCTC 8594 to identify presence/absence of bacteriophage

Further investigation of differences between the two *C. sporogenes* PA 3679 strains was conducted by sequencing of the genome of *C. sporogenes* ATCC 7955 and aligning the resultant reads to the genome of *C. sporogenes* NCTC 8954. This

strategy however did not allow the identification of novel genetic regions present in the ATCC strain, however absent from the NCTC strain.

Visualisation of the alignment identified broad genomic similarities however with significant differences in those regions associated with the phage identified in Chapter 3 (Figure 6.9). Compared to the 10 phage regions identified in NCTC 8594 only two incomplete phage regions (Clostri_PA3679_φ01 Clostri_PA3679_φ07) were present in the ATCC 7955 strain.

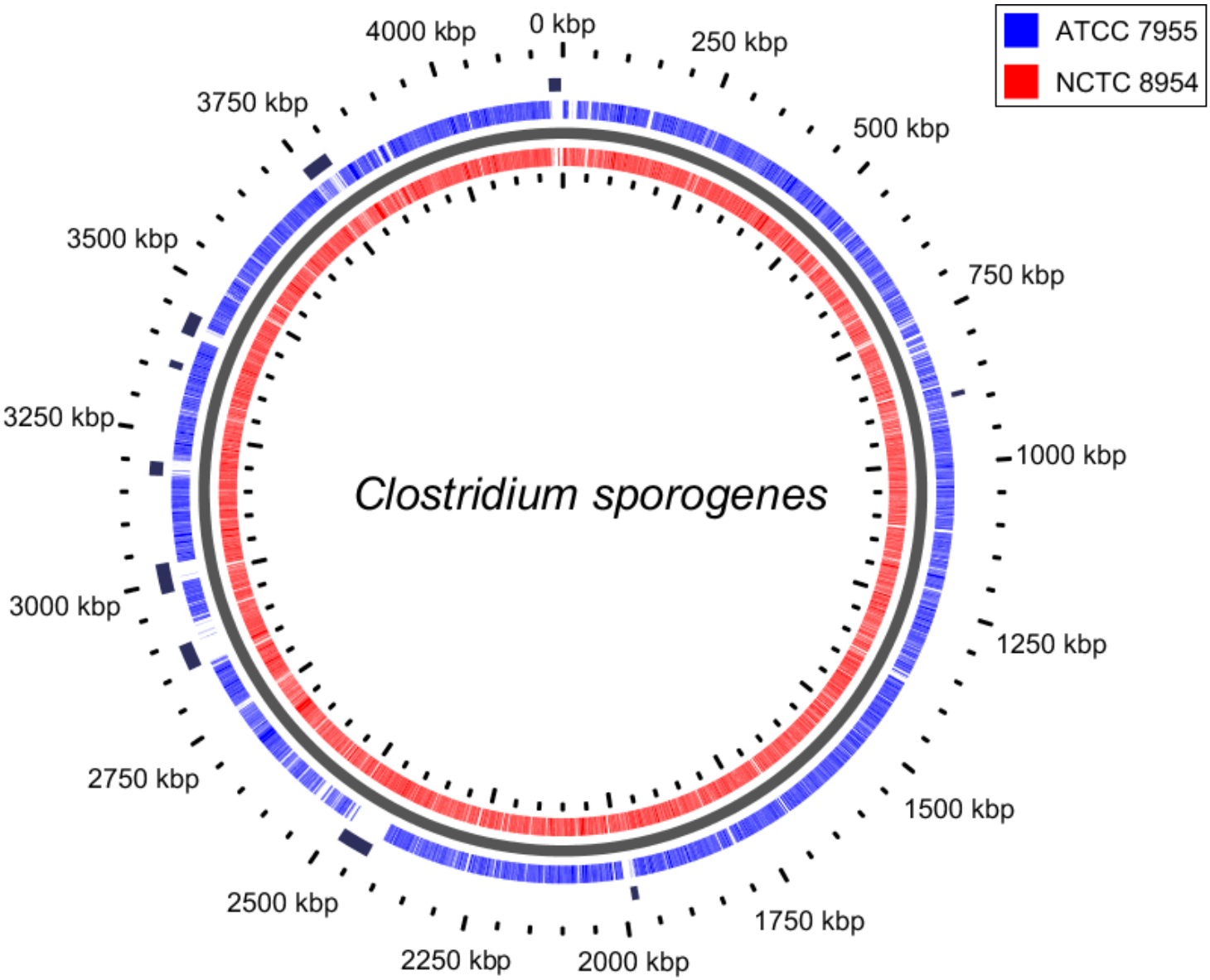

Figure 6.9 Alignment of Illumina MiSeq reads of *C. sporogenes* ATCC 7955 to the pseudomolecule of *C. sporogenes* NCTC 8594. Locations of phage regions identified by PHAST are represented by the dark blue bars on the outside ring.

Comparison between the toxin and germination regions in the NCTC isolate described previously (Chapter 3) and the ATCC isolate revealed similarity with the identification of some potential single nucleotide polymorphism in these regions based on the mapping. Due to the relatively low coverage these were not further investigated.

6.4 Discussion

Two strains of *C. sporogenes* PA 3679 sourced from different culture collections (ATCC 7955, NCTC 8594) were investigated to identify potential factors associated with variation in heat resistance. Significant difference in spore heat resistance was observed between strains both using recovery media and as assessed by flow cytometry. The ultrastructures of dormant spores revealed a decreased spore core and increased exosporium for the more heat resistant ATCC strain compared to the NCTC strain. Genomic comparison between the strains also revealed significant variation in the number of phage previously described in Chapter 3.

The heat resistance properties of spores produced from *C. sporogenes* ATCC 7955 were relatively similar to the reported industry standard $D_{121.1} = 0.21$ min for Group I *C. botulinum*, consistent with previous findings of this thesis that with the notable exception in BoNT associated genes *C. sporogenes* are genetically indistinct from Group I *C. botulinum* (see Chapter 4), and as such are unlikely to exhibit significantly different phenotypic characteristics. The NCTC spore crop however did exhibit abnormally low heat resistance compared to reported *C. sporogenes* and Group I *C. botulinum* D-values (Diao et al., 2014).

Further investigation in the differences between spore populations derived from each strain using flow cytometry revealed significant differences in the rate at which spores became permeable to fluorogenic dyes during a thermal process. The significantly faster rate at which NCTC spores underwent transition from a dye impermeable to dye permeable state was indicative of a higher susceptibility of the NCTC spore structure to heat. Staining of spores with fluorophores including SYTO 9 and PI has been associated with permeation of the spore inner membrane and hydrolysis of the spore cortex (Setlow, 2003). The differences in the permeability between spore crops

indicates that inactivation of NCTC spores is most likely associated with loss of spore structure rather than inactivation of proteins involved in spore outgrowth such as cortex lytic enzymes.

Investigation of the spore structure via TEM revealed a stark contrast between the two spore crops in the size of the exosporium. The ATCC spores retained a distinctly more voluminous exosporium compared to the NCTC spores. The origin of the closer juxtaposition of the exosporium to the outer coat of the dormant NCTC spores is unknown. Similar positioning has been observed in *Bacillus* spores grown in a glucose rich medium and associated with the down regulation of essential coat assembly gene *cotγ* (Aronson et al., 2014). The significant differences in the exosporium reveal at least one potential source of heat resistance variability of spore crops. As the outermost layer of spores, the exosporium is central to the response of a spore to the external environment. One possible hypothesis based on the exosporium size measurements, coupled with the flow cytometric data, is that the variation in exosporium size may alter the relative response of spores to external environmental triggers resulting in a more rapid inactivation of germinated spores. Alterations in the presence and size of the exosporium have been previously correlated with increased spore hydrophobicity (Takubo et al., 1988) and have also been associated with decreased germinant response and spore resistance to lytic enzymes.

Differences observed in core and exosporium size may be attributable to the prevention of full maturation of the NCTC strain spores. Both core dehydration and development of the exosporium typically occur late during the sporulation process (Stevenson and Vaughn, 1972). The underdevelopment of these aspects of the spore may be attributable to premature release of the spore possibly by lysis of the mother cell. This was not readily observed through phase contrast microscopy of the sporulating cultures, however it was not able to be completely discounted as a potential source of variation between the strains.

Due to the relatively low coverage sequencing the genome comparison between strains was limited to identifying the absence of large genomic regions (> 3 kbp) compared to the NCTC 8594 genome. These differences were predominately associated with regions associated with prophage. Both strains have been used in multiple studies, with the *C. sporogenes* NCTC 8594 isolate having been used

consistently for the last sixty years (Roberts and Hoover, 1966; Ashworth and Spencer, 1972; Poxton, 1984; Aranda et al., 1997; Bull et al., 2009) and the ATCC 7955 even longer (Curran et al., 1943; Lund et al., 1957; Savani et al., 1978; Wiencek et al., 1990; Ramaswamy and Shao, 2010). The confirmed significant genetic differences between the strain suggest that conclusions based on *C. sporogenes* PA 3679 from specific collections should be assessed with care. Furthermore, broader implications regarding generalisations of species specific heat resistance values must be viewed with a degree of scepticism and risk management strategies employing these values would benefit from validation with multiple strains.

In order to better validate these findings, deeper sequencing of the *C. sporogenes* ATCC 7955 strain may confirm whether differences observed are solely based on bacteriophage differences.

6.5 Concluding remarks

Significant differences were identified in the heat resistance characteristics of the spores of two *C. sporogenes* strains both identified as "PA 3679". Flow cytometric and TEM investigation of the spores identified that the heat resistance differences could at least partially be attributed to variation in the spore structure. Studies of the ultrastructural differences in spores revealed that the spores with characteristics of mature spores including a significant larger exosporium and dehydrated core exhibited greater heat resistance.

Genetic comparison of the two strains revealed stark contrast in specific regions indicating that despite both being identified as "PA 3679" by their respective culture collections the strains exhibited significant genetic differences, predominately in regions associated with bacteriophage. The differentiation of these strains by their culture collection designation (ATCC 7955 and NCTC 8594), rather than their historical designation as "PA 3679" is highly recommended and previous meta-analysis studies conducted using data derived from both strains may need to be re-examined. Furthermore, this study highlights the previous observations of this thesis that bacteriophage are a major determinant of genetic diversity within the *C. botulinum*/*C. sporogenes* group of organisms.

Chapter 7 – General Discussion and Concluding Remarks

7.1 Introduction

The studies reported within this thesis produced a number of significant research outcomes contributing to an enhanced understanding of the relationship between *C. sporogenes* and Group I *C. botulinum*, the genetic diversity of this group, and structural determinants contributing to the thermo-tolerance of their spores. The sequencing of the genome of *C. sporogenes* PA 3679 facilitated the identification of likely genetic contributors to the previously described heat resistance in this strain. Furthermore, the integration of this genome into a pan-genome of Group I *C. botulinum* highlighted the significant insights into the genetic diversity of this group of organisms and identified the importance of bacteriophage in this genetic diversity. Investigation of the heat inactivation kinetics of single spores of *C. sporogenes* described the impact of food systems formulation on the rate of inactivation of individual spores. Structural determinants correlating with the heat resistance properties of spores were observed, most markedly in exosporium size.

This chapter broadly summarises the discussions and major findings in previous chapters, whilst highlighting the importance of such outcomes to the aim of this research to further understand the genetic relationship between *C. sporogenes* and *C. botulinum*; and identifying the potential contributors to the heat resistance properties of spores.

7.2 Sequencing of the genome of *C. sporogenes* PA 3679 identified extensive prophage sequences and potential genetic contributors to increased heat resistance

The sequencing of the genome of *C. sporogenes* PA 3679 (Chapter 3) was conducted to assess specific genetic traits due to its wide spread usage as a heat resistant surrogate for heat resistant *C. botulinum*. The genome was observed to contain an extensive number of putative prophage regions. Ten prophage regions were identified representing the majority of strain specific DNA in the sequence, this is significantly more than the two prophage identified in *C. botulinum* ATCC 3502 (Sebaihia et al.,

2007). The presence of such a large number of prophage suggests a significant role in the niche adaptation of this strain and highlights the susceptibility of this strain to acquire exogenous genetic material from the environment.

A potential genetic contributor to heat resistance was observed in an atypical number of germination operons. The genome of *C. sporogenes* PA 3679 contained five polysistronic germination operons, in comparison to the three to four germination operons identified in Group I *C. botulinum* (Xiao et al., 2011). The structure of these germination operons was broadly consistent with those previously described in *Bacillus* spp., and *Clostridium* spp. (Ross and Abel-Santos, 2010). The structure of at least one of the germination operons described was consistent with the germination operon previously described in *C. sporogenes* identified using degenerate primers (Broussolle et al., 2002). The greater number of germination operons suggests the potential for spores of *C. sporogenes* to exhibit either a higher specificity and/or sensitivity to a range of germinants. Previous studies have been unable to identify differences in germinant specificity between *C. sporogenes* and *C. botulinum* (Broussolle et al., 2002). Given the significance of these different germination operons, further research into these germination operons and variation in germination between these spores may provide significant opportunities for "germinate and exterminate" strategies to meet food safety objectives.

7.3 *C. sporogenes* strains are phylogenetically indistinguishable from Group I *C. botulinum* strains

Pan-genomic and MLST phylogenetic analyses were conducted to attempt to differentiate between non-toxigenic C. *sporogenes* and toxigenic Group I C. *botulinum* strains. On the basis of these analyses it was concluded that toxigenicity could not be related to the phylogenetic relationship of strains. The phylogenetic positioning revealed an inconsistent relationship between BoNT toxigenicity and phylogeny, as C. *botulinum* and C. *sporogenes* strains were often co-located within the same clade. This finding is consistent with the proposed lateral acquisition of BoNT in these strains. The incongruent phylogeny of BoNT genes and host has been described previously (Collins and East, 1998); however studies have not previously been conducted using MLST. The impact of this finding highlights that phylogenetic

methods alone are unlikely to provide suitable discrimination between toxigenic and non-toxigenic strains.

The most effective genetic method to determine the toxigenicity of a strain is the identification of BoNT encoding genes. However, the variability in genomic location of these genes coupled with the high degree of variation of the BoNT and ANTP CDS's make this a complex task. An alternative genetic approach to differentiate between toxigenic and non-toxigenic strains may be through the identification of other genetic markers. The pan-genomic analysis presented has provided a number of potential species specific candidate genes; 262 genes specific to *C. sporogenes* and 165 *C. botulinum* specific genes. These genes represent potential targets for future biomarkers, however due to the limited sample size of only two *C. sporogenes* genomes and ten *C. botulinum* genomes, these candidate genes must be viewed with some caution and reviewed as the number of available genomes increases.

7.4 Phylogenetic analysis provided potential mechanism of BoNT evolution and strain level differentiation

Another important finding from this study provides evidence toward the horizontal transfer of BoNT within Group I *C. botulinum*. The group of genes required for BoNT production in Group I *C. botulinum* are located either on the chromosome or on plasmids. The phylogenetic analysis of *C. sporogenes* found that all strains from this study were co-located in a similar clade to Group I C. *botulinum* with plasmid encoded BoNT (Figure 4.2).

The co-location of non-toxigenic *C. sporogenes* with plasmid borne strains of Group I C. *botulinum* provides strong evidence toward plasmid mediated BoNT transfer within this clade. This is in contrast to Group 3 C. *botulinum* where BoNT is encoded on bacteriophage (Sakaguchi et al., 2005). Significantly the strains of *C. botulinum* that were most closely related to *C. sporogenes* were associated with infant botulism rather than foodborne cases of botulism. The separation of clades of organisms associated with foodborne and infant botulism indicates that the causative agents of these forms of botulism may differ. Increased surveillance of both clinical and environmental isolates coupled with whole genome studies will provide significantly greater insights, both into the epidemiological relationship of *C. botulinum* strains

associated with specific cases of botulism and the genetic flow of BoNT through this group of organisms.

7.5 Bacteriophage represent a major driver in the genetic diversity of *C. sporogenes* and Group I *C. botulinum*

Investigation into the pan-genome of *C. sporogenes* and Group I *C. botulinum* identified bacteriophage as the major driver in the genetic diversity within this group. Of the ten Group I *C. botulinum* strains and two *C. sporogenes* strains the vast majority of strain specific genes were associated with regions identified as putative prophage. This result indicates the occurrence of significant levels of acquired genetic material from the environment within this group of organisms. The number of bacteriophage associated within *C. sporogenes* and Group I *C. botulinum* suggest that strain level differentiation using phage typing may have some application. However, differences in toxigenicity by phage alone may present some difficult as previously demonstrated by Kiritani et al., (1973) who were unable to differentiate between *C. botulinum* and *C. sporogenes* strains based on susceptibility to induced lysins. Further research on the bacteriophage present in *C. sporogenes* and *C. botulinum* may also provide insight into the relationship between bacteriophage and phenotypic traits including, sporulation, germination and virulence.

7.6 Spore structure and structural integrity alterations contribute to the sensitivity of spores to heat treatments

Both the ultrastructure of spores in a resting state (as observed via TEM) and the rate at which spores transition from this resting state were identified as important factors in spore survival. The concentric ring structure of bacterial spores is essential for their dormancy and resistance to environmental stresses. Significant differences in the ultrastructure between heat resistant and heat sensitive *C. sporogenes* spores were identified. This variation in spore structure was most significant in the core, cortex and exosporium. A dehydrated core and an increased exosporium volume were found to correlate with increased resistance to heat. Both of these size alterations occur in

the later stages of spore formation indicative of more mature spores. These results differ from those of Mah et al., (2008a) who identified the cortex size as the primary structural determinant in spore heat resistance. The differences between these findings may be attributed to the different spore crop preparations and strain variability.

The importance of spore structural integrity was further highlighted using flow cytometry in the presence of salt as an inhibitor of spore germination. Using fluorogenic dyes the structural integrity of spores was assessed during a heat treatment. Spores that were able to maintain their structural integrity to prevent the ingress of fluorophores were shown to survive significantly longer than spores that were unable to maintain impermeability to the dyes.

7.7 Spore biovariability should be taken into account when describing the response of spores to heat

The heat resistance properties of *C. sporogenes* spores were extensively studied at the single cell level, both by TEM and flow cytometry. Significant biovariability was observed within both single spore crops and between strains, highlighting this as an important component contributing to spore heat resistance.

An example of clonal level biovariability was in exosporium size prior to a heat treatment. That such biovariability was observed after removal of vegetative cells and enrichment of spores using a density gradient indicates that such variability is even greater in naturally occurring spore populations. Another example of clonal level biovariability was the rate in which individual spores within a spore crop transitioned from a dye permeable to a dye impermeable state as assessed by flow cytometry.

Strain level variability was observed in the genomic content and spore ultrastructure of the *C. sporogenes* strains examined within this thesis. In particular, the differences between the ATCC strain *of C. sporogenes* PA 3679 and the NCTC strain revealed significant differences in bacteriophage and heat resistant properties (Chapter 6).

The ramifications of such variability in predictive food microbiology when extrapolating results from experiments can be quite severe, for example the difference in time required to achieve a 12-log reduction at 121.1°C for spores based on the heat resistance values derived from the ATCC strain of *C. sporogenes* from this study would be ~3.7 minutes vs. ~0.6 minutes for the NCTC strain of *C. sporogenes*. This

biovariability suggests that significant care must be taken in deriving appropriate processing parameters and in challenge studies. It also suggests that selection of a single strain, as a target organism may not represent the most appropriate approach to designing food safety strategies. Multi-strain challenge studies are likely to provide a significantly more robust approach for risk based food safety management.

7.8 Recommendations for future research

A number of opportunities for further research are presented in order to improve our understanding of the genetics of *Clostridium* spp. both in the transfer of toxigenicity within strains and heat resistance of spores. Primarily, increasing the quantity of strains will provide significantly improved understanding of the diversity within Group I *C. botulinum*, since currently all strains which have been investigated have been selected due to association with botulism. This selection bias influences our understanding of the genetics of this species. A sampling of environmental isolates may provide significantly more information and also identify geographical, environmental and epidemiological variation within this group of organisms.

With regard to the heat resistance of spores, the recognised importance of germination potential may provide significance to improved processing of foods. The genomic approach employed to identify germination operons is the first step in identifying the potential mechanism of spore germination in *C. sporogenes*. The construction of knockout mutants using increasingly available genetic manipulation techniques such as Clostron may provide insights into the function of the germinant receptors encoded by these operons.

Transcriptomic and proteomic studies in particular during the sporulation and germination processes may significantly enhance the understanding of the role of these germination operons and variation in spore structure. A common issue involved in these studies remains the heterogeneity of spore cultures and cells in differing states; flow cytometry coupled with fluorescence activated cell sorting is one potential approach that may allow for the application of such techniques.

Finally, the development of a mechanistic model for spore inactivation integrating the intrinsic variability associated with spores would be recommended. The application of more complex models of spore inactivation may provide significant improvements in the establishment of less conservative thermal process parameters. Probability models derived using Bayesian statistics are one possible avenue worth future exploration.

7.9 Concluding remarks

In this series of studies the genome of *C. sporogenes* PA 3679 was generated and employed to establish potential genetic contributors to the heat resistance of this strain and describe the close phylogenetic relationship to Group I *C. botulinum*. The genome consisting of 4111 genes was characterised by an abundance of bacteriophage and a high degree of synteny to reference *C. botulinum* genomes.

Construction of the pan-genome of *C sporogenes* and Group I *C. botulinum* showed a high degree of genetic diversity of this group of organisms with over 8799 genes identified in the pan-genome and a core genome consisting of 1590 genes. The diversity exhibited by strains was heavily influenced by the presence of prophage highlighting the presence of laterally acquired material, representing possible mechanisms for strain niche adaption.

A phylogenetic analysis of toxigenic and non-toxigenic strains provided further evidence for evolution of BoNT within Group I *C. botulinum* with conjugative plasmid exchange a likely route. Furthermore, the pan-genome was able to identify potential candidates for the differentiation of *C. sporogenes* and Group I *C. botulinum*.

Flow cytometry in combination with SYTO 9 and PI was able to discriminate between groups of spores exhibiting differing physiological states, highlighting the importance of spore state and heterogeneity in thermal process calculations. A mechanistic model describing the thermal inactivation of spores in the presence of NaCl was developed. Both an antagonistic and protective effect of NaCl was observed for heat treated spores. The presence of NaCl was observed to decrease the initial rate of spore inactivation by heat due to the loss of dormancy, however further enhanced the onset of membrane permeability of spores that were no longer dormant.

These studies provide enhanced knowledge of the genetics of *C. sporogenes* PA 3679 and the pan-genome of Group I *C. botulinum* and *C. sporogenes*. It has been identified that whilst *C. sporogenes* is a suitable surrogate for Group I *C. botulinum* and shares a high degree of genetic similarity to Group I *C. botulinum* it also exhibits significant differences. Such genetic variability suggests that future studies of *C. sporogenes* PA 3679 as a surrogate should ensure that these differences do not contribute to variable phenotypic traits under the multitude of environmental stressors encountered in food systems. Furthermore, this study identified significant diversity within the pan-genome of *C. botulinum* and *C. sporogenes* suggesting that a "typical" *C. botulinum* does not exist and as such phenotypic screening of strains for specific resistances remains the most appropriate method in the selection of a surrogate organism.

Chapter 8 - References

Abbott, J. C., Aanensen, D. M., Rutherford, K., Butcher, S., & Spratt, B. G. (2005). WebACT—an online companion for the Artemis Comparison Tool. *Bioinformatics*, *21*(18), 3665-3666.

Adams, D. M. (1978). Heat injury of bacterial spores. *Advances in Applied Microbiology*, *23*, 245-261.

Albert, I., & Mafart, P. (2005). A modified Weibull model for bacterial inactivation. *International Journal of Food Microbiology*, *100*(1–3), 197–211.

Alberto, F., Botella, L., Carlin, F., Nguyen-the, C., & Broussolle, V. (2005). The *Clostridium botulinum* GerAB germination protein is located in the inner membrane of spores. *FEMS Microbiology Letters*, *253*(2), 231–235.

Alcaraz, L. D., Moreno-Hagelsieb, G., Eguiarte, L. E., Souza, V., Herrera-Estrella, L., & Olmedo, G. (2010). Understanding the evolutionary relationships and major traits of *Bacillus* through comparative genomics. *BMC Genomics*, *11*(1), 332.

Anderson, W. A. Non-log-linear thermal inactivation of *Clostridium botulinum* spores, Ph.D. thesis, University of Reading, 1997.

Aranda, E., Rodriguez, M. M., Asensio, M. A., & Cordoba, J. J. (1997). Detection of *Clostridium botulinum* types A, B, E and F in foods by PCR and DNA probe. *Letters in Applied Microbiology*, *25*(3), 186-190.

Aronson, A., Goodman, B., & Smith, Z. (2014). The regulated synthesis of a *Bacillus anthracis* spore coat protein that affects spore surface properties. *Journal of Applied Microbiology*, 116, 1241-1249.

Ashworth, J., & Spencer, R. (1972). The Perigo effect in pork. *International Journal of Food Science & Technology*, *7*(2), 111-124.

van Asselt, E. D., & Zwietering, M. H. (2006). A systematic approach to determine global thermal inactivation parameters for various food pathogens. *International Journal of Food Microbiology*, *107*(1), 73–82.

Atrih, A., & Foster, S. J. (2002). Bacterial endospores the ultimate survivors. *International Dairy Journal*, *12*(2–3), 217–223.

Augustin, J. A., & Pflug, I. J. (1967). Recovery patterns of spores of putrefactive anaerobe 3679 in various subculture media after heat treatment. *Applied Microbiology*, *15*(2), 266–276.

Avery, S. V. (2006). Microbial cell individuality and the underlying sources of heterogeneity. *Nature Reviews Microbiology*, *4*(8), 577–587.

Ball, D. A., Taylor, R., Todd, S. J., Redmond, C., Couture-Tosi, E., Sylvestre, P., … Bullough, P. A. (2008). Structure of the exosporium and sublayers of spores of the *Bacillus cereus* family revealed by electron crystallography. *Molecular Microbiology*, *68*(4), 947–958.

Barash, J. R., & Arnon, S. S. (2013). A Novel Strain of *Clostridium botulinum* That Produces Type B and Type H Botulinum Toxins. *Journal of Infectious Diseases*, jit449.

Bassi, D., Cappa, F., & Cocconcelli, P. S. (2013). Array-based transcriptional analysis of *Clostridium sporogenes* UC9000 during germination, cell outgrowth and vegetative life. *Food Microbiology*, *33*(1), 11–23.

Beaman, T. C., Greenamyre, J. T., Corner, T. R., Pankratz, H. S., & Gerhardt, P. (1982). Bacterial spore heat resistance correlated with water content, wet density, and protoplast/sporoplast volume ratio. *Journal of Bacteriology*, *150*(2), 870-877.

Beaman, T. C., & Gerhardt, P. (1986). Heat resistance of bacterial spores correlated with protoplast dehydration, mineralization, and thermal adaptation. *Applied and Environmental Microbiology*, *52*(6), 1242–1246.

den Besten, H. M. W., Ingham, C. J., van Hylckama Vlieg, J. E. T., Beerthuyzen, M. M., Zwietering, M. H., & Abee, T. (2007). Quantitative analysis of population heterogeneity of the adaptive salt stress response and growth capacity of *Bacillus cereus* ATCC 14579. *Applied and Environmental Microbiology*, *73*(15), 4797–4804.

Broussolle, V., Alberto, F., Shearman, C. A., Mason, D. R., Botella, L., Nguyen-The, C., … Carlin, F. (2002). Molecular and physiological characterisation of spore germination in *Clostridium botulinum* and *C. sporogenes*. *Anaerobe*, *8*(3), 89–100.

Brown, J. L., Tran-Dinh, N., & Chapman, B. (2012). *Clostridium sporogenes* PA 3679 and Its uses in the derivation of thermal processing schedules for low-acid shelf-stable foods and as a research model for proteolytic *Clostridium botulinum*. *Journal of Food Protection*, *75*(4), 779–792.

Budde, B. B., & Rasch, M. (2001). A comparative study on the use of flow cytometry and colony forming units for assessment of the antibacterial effect of bacteriocins. *International Journal of Food Microbiology*, *63*(1), 65-72.

Bull, M. K., Olivier, S. A., Diepenbeek, R. J. van, Kormelink, F., & Chapman, B. (2009). Synergistic inactivation of spores of proteolytic *Clostridium botulinum* strains by high pressure and heat is strain and product dependent. *Applied and Environmental Microbiology*, *75*(2), 434–445.

Cabrera-Martinez, R. M., Tovar-Rojo, F., Vepachedu, V. R., & Setlow, P. (2003). Effects of overexpression of nutrient receptors on germination of spores of *Bacillus subtilis*. *Journal of Bacteriology*, *185*(8), 2457-2464.

Carr, K. A., Janes, B. K., & Hanna, P. C. (2010). Role of the *gerP* operon in germination and outgrowth of *Bacillus anthracis* spores. *PLoS ONE, 5*(2), e9128.

Carter, A. T., Mason, D. R., Grant, K. A., Franciosa, G., Aureli, P., & Peck, M. W. (2010). Further characterization of proteolytic *Clostridium botulinum* type A5 reveals that neurotoxin formation is unaffected by loss of the *cntR* (*botR*) promoter sigma factor binding site. *Journal of Clinical Microbiology, 48*(3), 1012–1013.

Carter, A. T., Paul, C. J., Mason, D. R., Twine, S. M., Alston, M. J., Logan, S. M., … Peck, M. W. (2009). Independent evolution of neurotoxin and flagellar genetic loci in proteolytic *Clostridium botulinum*. *BMC Genomics, 10*(1), 115.

Cerf, O. (1977). Tailing of survival curves of bacterial spores. *Journal of Applied Microbiology, 42*(1), 1–19.

Chen, F., Kuziemko, G. M., & Stevens, R. C. (1998). Biophysical characterization of the stability of the 150-kilodalton botulinum toxin, the nontoxic component, and the 900-kilodalton botulinum toxin complex species. *Infection and Immunity, 66*(6), 2420–2425.

Coil, D., Jospin, G., & Darling, A. E. (2014). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *arXiv preprint arXiv:1401.5130*.

Coleman, W. h., Zhang, P., Li, Y. -q., & Setlow, P. (2010). Mechanism of killing of spores of *Bacillus cereus* and *Bacillus megaterium* by wet heat. *Letters in Applied Microbiology*, *50*(5), 507–514.

Coleman, W. H., & Setlow, P. (2009). Analysis of damage due to moist heat treatment of spores of *Bacillus subtilis*. *Journal of Applied Microbiology*, *106*(5), 1600-1607.

Coleman, W. H., Chen, D., Li, Y., Cowan, A. E., & Setlow, P. (2007). How moist heat kills spores of *Bacillus subtilis*. *Journal of Bacteriology*, *189*(23), 8458–8466.

Collins, M.D., & East, A.K. (1998). Phylogeny and taxonomy of the food-borne pathogen *Clostridium botulinum* and its neurotoxins. *Journal of Applied Microbiology, 84*(1), 5–17.

Coote, P. J., Billon, C. P., Pennell, S., McClure, P. J., Ferdinando, D. P., & Cole, M. B. (1995). The use of confocal scanning laser microscopy (CSLM) to study the germination of individual spores of *Bacillus cereus*. *Journal of Microbiological Methods*, *21*(2), 193-208.

Corradini, M. G., Normand, M. D., & Peleg, M. (2010). Stochastic and deterministic model of microbial heat inactivation. *Journal of Food Science*, *75*(2), R59-R70.

Cowan, A. E., Olivastro, E. M., Koppel, D. E., Loshon, C. A., Setlow, B., & Setlow, P. (2004). Lipids in the inner membrane of dormant spores of *Bacillus* species are largely immobile. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(20), 7733–7738.

Cronin, U. P., & Wilkinson, M. G. (2007). The use of flow cytometry to study the germination of *Bacillus cereus* endospores. *Cytometry Part A*, *71*(3), 143-153.

Curran, H. R., Brunstetter, B. C., & Myers, A. T. (1943). Spectrochemical analysis of vegetative cells and spores of bacteria. *Journal of Bacteriology*,*45*(5), 485.

Curran, H. R., & Evans, F. R. (1945). Heat activation inducing germination in the spores of thermotolerant and thermophilic aerobic bacteria. *Journal of Bacteriology*, *49*(4), 335–346.

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*, *5*(6), e11147.

Davey, H. M., & Hexley, P. (2011). Red but not dead? Membranes of stressed *Saccharomyces cerevisiae* are permeable to propidium iodide. *Environmental Microbiology*, *13*(1), 163-171.

Davey, H. M., & Kell, D. B. (1996). Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. *Microbiological Reviews*, *60*(4), 641-696.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., ... & Gascuel, O. (2008). Phylogeny. fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, *36*(suppl 2), W465-W469.

Diao, M. M., André, S., & Membré, J. M. (2014). Meta-analysis of D-values of proteolytic *Clostridium botulinum* and its surrogate strain *Clostridium sporogenes* PA 3679. *International Journal of Food Microbiology*, 174, 23-30.

Dover, N., Barash, J. R., Hill, K. K., Xie, G., & Arnon, S. S. (2014). Molecular characterization of a novel botulinum neurotoxin type H gene. *Journal of Infectious Diseases*, *209*(2), 192-202.

Dover, N., Barash, J. R., Hill, K. K., Davenport, K. W., Teshima, H., Xie, G., & Arnon, S. S. (2013). *Clostridium botulinum* Strain Af84 contains three neurotoxin gene clusters: bont/A2, bont/F4 and bont/F5. *PLoS ONE*, *8*(4), e61205.

Driks, A. (1999). *Bacillus subtilis* spore coat. *Microbiology and Molecular Biology Reviews*, *63*(1), 1–20.

Driks, A. (2007). Surface appendages of bacterial spores. *Molecular Microbiology*, *63*(3), 623-625.

Duncan, C. L., & Foster, E. M. (1968). Effect of sodium nitrite, sodium chloride, and sodium nitrate on germination and outgrowth of anaerobic spores. *Applied Microbiology*, *16*(2), 406-411.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*(1), 113.

Eijlander, R. T., Abee, T., & Kuipers, O. P. (2011). Bacterial spores in food: how phenotypic variability complicates prediction of spore properties and bacterial behaviour. *Current Opinion in Biotechnology*, *22*(2), 180–186.

Eklund, M. W., Poysky, F. T., & Reed, S. M. (1972). Bacteriophage and the toxigenicity of *Clostridium botulinum* Type D. *Nature*, *235*(53), 16–17.

English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., ... & Gibbs, R. A. (2012). Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PloS one*, *7*(11), e47768.

Erbguth, F. J. (2004). Historical notes on botulism, *Clostridium botulinum*, botulinum toxin, and the idea of the therapeutic use of the toxin. *Movement Disorders*, *19*(S8), S2–S6.

van Ermengem, E. P. (1897). Ueber einen neuen anaeroben *Bacillus* und seine Beziehungen zum Botulismus. *Medical Microbiology and Immunology*, *26*(1), 1-56.

Errington, J. (1993). *Bacillus subtilis* sporulation: regulation of gene expression and control of morphogenesis. *Microbiological Reviews*, *57*(1), 1–33.

Errington, J. (2003). Regulation of endospore formation in *Bacillus subtilis*. *Nature Reviews. Microbiology*, *1*(2), 117–126.

Fang, P.-K., Raphael, B. H., Maslanka, S. E., Cai, S., & Singh, B. R. (2010). Analysis of genomic differences among *Clostridium botulinum* type A1 strains. *BMC Genomics*, *11*(1), 725.

Flowers, R. S., & Adams, D. M. (1976). Spore membrane(s) as the site of damage within heated *Clostridium perfringens* spores. *Journal of Bacteriology*, *125*(2), 429–434.

Foegeding, P. M., & Busta, F. F. (1983). Hypochlorite injury of *Clostridium botulinum* spores alters germination responses. *Applied and Environmental Microbiology*, *45*(4), 1360–1368.

Franciosa, G., Floridi, F., Maugliani, A., & Aureli, P. (2004). Differentiation of the Gene Clusters Encoding Botulinum Neurotoxin Type A Complexes in

Clostridium botulinum Type A, Ab, and A(B) Strains. *Applied and Environmental Microbiology*, *70*(12), 7192–7199.

Franciosa, G., Maugliani, A., Scalfaro, C., & Aureli, P. (2009). Evidence that plasmid-borne botulinum neurotoxin type B genes are widespread among *Clostridium botulinum* serotype B strains. *PLoS ONE*, *4*(3), e4829.

Ferencko, L., & Rotman, B. (2010). Constructing fluorogenic *Bacillus* spores (F-spores) via hydrophobic decoration of coat proteins. *PloS one*, *5*(2), e9283.

FSANZ, 2007. Food Surveillance News - Spring 2007. Available at: http://www.foodstandards.gov.au/science/monitoring/surveillance/pages/news/spring2007.aspx [Accessed December 17, 2013].

Fujinaga, Y., Sugawara, Y., & Matsumura, T. (2013). Uptake of botulinum neurotoxin in the intestine. *Current Topics in Microbiology and Immunology*, *364*, 45–59

Fujioka, R. S., & Frank, H. A. (1966). Nutritional requirements for germination, outgrowth, and vegetative growth of putrefactive anaerobe 3679 in a chemically defined medium. *Journal of Bacteriology*, *92*(5), 1515–1520.

Gerhardt, P., & Marquis, R. E. (1989). Spore thermoresistance mechanisms. In: Smith I, Slepecky R A, Setlow P, (Eds.), *Regulation of prokaryotic development*. Washington, D.C: American Society for Microbiology. pp. 43–63.

Ghosh, S., & Setlow, P. (2009). Isolation and Characterization of superdormant spores of *Bacillus* species. *Journal of Bacteriology*, *191*(6), 1787–1797.

Goecks, J., Li, K., Clements, D., & Taylor, J. (2011). The Galaxy Track Browser: transforming the genome browser from visualization tool to analysis tool. In *Biological Data Visualization (BioVis), 2011 IEEE Symposium on* (pp. 39-46). IEEE.

Gordienko, E. N., Kazanov, M. D., & Gelfand, M. S. (2013). Evolution of pan-genomes of *Escherichia coli*, *Shigella* spp., and *Salmonella enterica*. *Journal of Bacteriology*, *195*(12), 2786-2792.

Gorman, S. P., Hutchinson, E. P., Scott, E. M., & McDermott, L. M. (1983). Death, injury and revival of chemically treated *Bacillus subtilis* spores. *Journal of Applied Bacteriology*, *54*(1), 91–99.

Gould, G. W., & Dring, G. J. (1975). Heat resistance of bacterial endospores and concept of an expanded osmoregulatory cortex. *Nature*, *258*(5534), 402–405.

Gould, G. W., & Dring, G. J. (1974). Mechanisms of spore heat resistance. *Advances in Microbial. Physiology*, *11*, 137–164.

Grischy, R. O., Speck, R. V., & Adams, D. M. (1983). New media for enumeration and detection of *Clostridium sporogenes* (PA3679) Spores. *Journal of Food Science*, *48*(5), 1466-1469.

Grossman, A. D. (1995). Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Annual Review of Genetics*, *29*, 477–508.

Gu, S., & Jin, R. (2013). Assembly and function of the botulinum neurotoxin progenitor complex. *Current Topics in Microbiology and Immunology*, *364*, 21–44.

Hachisuka, Y., Kozuka, S., & Tsujikawa, M. (1984). Exosporia and appendages of spores of *Bacillus* species. *Microbiology and Immunology*, *28*(5), 619–624.

Hall, B. G., Ehrlich, G. D., & Hu, F. Z. (2010). Pan-genome analysis provides much higher strain typing resolution than multi-locus sequence typing.*Microbiology*, *156*(4), 1060-1068.

Hall, J. D., McCroskey, L. M., Pincomb, B. J., & Hatheway, C. L. (1985). Isolation of an organism resembling *Clostridium barati* which produces type F botulinal toxin from an infant with botulism. *Journal of Clinical Microbiology*, *21*(4), 654–655.

Hamouda, T., Shih, A. Y., & Baker, J. R. (2002). A rapid staining technique for the detection of the initiation of germination of bacterial spores. *Letters in Applied Microbiology*, *34*(2), 86-90.

Haraldsen, J. D., & Sonenshein, A. L. (2003). Efficient sporulation in *Clostridium difficile* requires disruption of the σK gene. *Molecular Microbiology*, *48*(3), 811-821.

Hassan, K. A., Tetu, S. G., Elbourne, L. D. H., Johnson, E. A., & Paulsen, I. T. (2013). Genome sequence of the Group III *Clostridium botulinum* strain Eklund-C. *Genome Announcements*, *1*(2), e00044–13.

Hauschild, A. 1993. Epidemiology of Human Foodborne Botulism. In: Hauschild, (Ed.).*Clostridium botulinum : Ecology and Control in Foods*, New York: M. Dekker.

Hauschild, A. H., & Dodds, K. L. (1993). *Clostridium botulinum: Ecology and Control in foods*. CRC Press.

Hill, K. K., & Smith, T. J. (2013). Genetic diversity within *Clostridium botulinum* serotypes, botulinum neurotoxin gene clusters and toxin subtypes. In A. Rummel & T. Binz (Eds.), *Botulinum Neurotoxins* (pp. 1–20). Springer Berlin Heidelberg.

Hill, K. K., Xie, G., Foley, B. T., Smith, T. J., Munk, A. C., Bruce, D., … Detter, J. C. (2009). Recombination and insertion events involving the botulinum neurotoxin

complex genes in *Clostridium botulinum* types A, B, E and F and *Clostridium butyricum* type E strains. *BMC Biology*, *7*(1), 66.

Hill, K. K., Smith, T. J., Helma, C. H., Ticknor, L. O., Foley, B. T., Svensson, R. T., ... & Marks, J. D. (2007). Genetic diversity among botulinum neurotoxin-producing clostridial strains. *Journal of Bacteriology*, *189*(3), 818-832.

Hodgkiss, W., Ordal, Z. J., & Cann, D. C. (1967). The morphology and ultrastructure of the spore and exosporium of some *Clostridium* species. *Journal of General Microbiology, 47*(2), 213–225.

Hornstra, L. M., Ter Beek, A., Smelt, J. P., Kallemeijn, W. W., & Brul, S. (2009). On the origin of heterogeneity in (preservation) resistance of *Bacillus* spores: Input for a "systems" analysis approach of bacterial spore outgrowth. *International Journal of Food Microbiology*, *134*(1–2), 9–15.

Hudson, K. D., Corfe, B. M., Kemp, E. H., Feavers, I. M., Coote, P. J., & Moir, A. (2001). Localization of GerAA and GerAC germination proteins in the *Bacillus subtilis* spore. *Journal of Bacteriology, 183*(14), 4317–4322.

Hurst, A. (1977). Bacterial injury: a review. *Canadian Journal of Microbiology*, *23*(8), 935–944.

Hutson, R. A., Thompson, D. E., Lawson, P. A., Schocken-Itturino, R. P., Böttger, E. C., & Collins, M. D. (1993). Genetic interrelationships of proteolytic *Clostridium botulinum* types A, B, and F and other members of the *Clostridium botulinum* complex as revealed by small-subunit rRNA gene sequences. *Antonie van Leeuwenhoek, 64*(3-4), 273–283.

Hutton, M. T., Koskinen, M. A., & Hanlin, J. H. (1991). Interacting effects of pH and NaCl on heat resistance of bacterial spores. *Journal of Food Science*, *56*(3), 821-822.

Hwang, H. J., Lee, J. C., Yamamoto, Y., Sarker, M. R., Tsuchiya, T., & Oguma, K. (2007). Identification of structural genes for *Clostridium botulinum* type C neurotoxin-converting phage particles. *FEMS Microbiology Letters*,*270*(1), 82-89.

Hyytiä, E., Björkroth, J., Hielm, S., & Korkeala, H. (1999). Characterisation of *Clostridium botulinum* groups I and II by randomly amplified polymorphic DNA analysis and repetitive element sequence-based PCR. *International Journal of Food Microbiology*, *48*(3), 179–189.

Inoue, K., & Iida, H. (1970). Conversion of Toxigenicity in *Clostridium botulinum* Type C. *Japanese Journal of Microbiology*, *14*(1), 87–89.

Jack, D. L., Paulsen, I. T., & Saier, M. H. (2000). The amino acid/polyamine/organocation (APC) superfamily of transporters specific for amino acids, polyamines and organocations. *Microbiology*, *146*(8), 1797–1814.

Jacobsen, A., Hendriksen, R. S., Aaresturp, F. M., Ussery, D. W., & Friis, C. (2011). The *Salmonella enterica* pan-genome. *Microbial ecology*, *62*(3), 487-504.

Jacobson, M. J., Lin, G., Raphael, B., Andreadis, J., & Johnson, E. A. (2008). Analysis of neurotoxin cluster genes in *Clostridium botulinum* strains producing botulinum neurotoxin serotype A subtypes. *Applied and Environmental Microbiology*, *74*(9), 2778–2786.

JCICSB: Judicial Commission of the International Committee on Systematic Bacteriology. (1999). Rejection of *Clostridium putrificum* and conservation of *Clostridium botulinum* and *Clostridium sporogenes*-Opinion 69. *International Journal of Systematic Bacteriology*, *49 Pt 1*, 339.

Johnson, E. A., & Bradshaw, M. (2001). *Clostridium botulinum* and its neurotoxins: a metabolic and cellular perspective. *Toxicon: Official Journal of the International Society on Toxinology*, *39*(11), 1703–1722.

Johnson, E.A. & Montecucco, C., 2008. Chapter 11 Botulism. In *Neuromuscular Junction Disorders*. Elsevier, pp. 333-368.

Jones, S. W., Paredes, C. J., Tracy, B., Cheng, N., Sillers, R., Senger, R. S., & Papoutsakis, E. T. (2008). The transcriptional program underlying the physiology of clostridial sporulation. *Genome Biology*, *9*(7), R114.

Keto-Timonen, R., Nevas, M., & Korkeala, H. (2005). Efficient DNA Fingerprinting of *Clostridium botulinum* Types A, B, E, and F by amplified fragment length polymorphism analysis. *Applied and Environmental Microbiology*, *71*(3), 1148–1154.

Kiritani, K., Mitsui, N., Nakamura, S., & Nishida, S. (1973). Numerical taxonomy of *Clostridium botulinum* and *Clostridium sporogenes* strains, and their susceptibilities to induced lysins and to mitomycin C. *Japanese Journal of Microbiology*, *17*(5), 361-372.

Koepke, R., Sobel, J., & Arnon, S. S. (2008). Global occurrence of infant botulism, 1976–2006. *Pediatrics*, *122*(1), e73-e82.

Kong, L., Zhang, P., Yu, J., Setlow, P., & Li, Y. Q. (2010). Monitoring the kinetics of uptake of a nucleic acid dye during the germination of single spores of *Bacillus* species. *Analytical Chemistry*, *82*(20), 8717-8724.

Kooiman, W. J., & Geers, J. M. (1975). Simple and accurate technique for the determination of heat resistance of bacterial spores. *Journal of Applied Bacteriology*, *38*(2), 185-189.

Kuenne, C., Billion, A., Mraheil, M. A., Strittmatter, A., Daniel, R., Goesmann, A., ... & Chakraborty, T. (2013). Reassessment of the *Listeria monocytogenes* pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. *BMC Genomics*, *14*(1), 47.

Kuzminov, A. (1999). Recombinational Repair of DNA damage in *Escherichia coli* and bacteriophage λ. *Microbiology and Molecular Biology Reviews*, *63*(4), 751-813.

Laflamme, C., Lavigne, S., Ho, J., & Duchaine, C. (2004). Assessment of bacterial endospore viability with fluorescent dyes. *Journal of Applied Microbiology*, *96*(4), 684-692.

Latreille, P., Norton, S., Goldman, B. S., Henkhaus, J., Miller, N., Barbazuk, B., ... & Slater, S. (2007). Optical mapping as a routine tool for bacterial genome sequence finishing. *BMC Genomics*, *8*(1), 321.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li, Y., Setlow, B., Setlow, P., & Hao, B. (2010). Crystal structure of the GerBC component of a *Bacillus subtilis* spore germinant receptor. *Journal of Molecular Biology*, *402*(1), 8–16.

Lin, W. J., & Johnson, E. A. (1995). Genome analysis of *Clostridium botulinum* type A by pulsed-field gel electrophoresis. *Applied and Environmental Microbiology*, *61*(12), 4441–4447.

Lindström, M., Hinderink, K., Somervuo, P., Kiviniemi, K., Nevas, M., Chen, Y., Korkeala, H. (2009a). Comparative genomic hybridization analysis of two predominant nordic Group I (proteolytic) *Clostridium botulinum* Type B clusters. *Applied and Environmental Microbiology*, *75*(9), 2643–2651.

Lindström, M., Fredriksson-Ahomaa, M. and Korkeala, H. (2009b). Molecular epidemiology of Group I and II *Clostridium botulinum* In: Brüggemann, H (Ed.). *Clostridia: molecular biology in the post-genomic era*, Wymondham: Caister Academic.

Lindström, M., Kiviniemi, K., & Korkeala, H. (2006). Hazard and control of group II (non-proteolytic) *Clostridium botulinum* in modern food processing. *International Journal of Food Microbiology*, *108*(1), 92–104.

Lund, A. J., Janssen, F. W., & Anderson, L. E. (1957). Effect of culture filtrates on sporogenesis in a species of *Clostridium*. *Journal of Bacteriology*, *74*(5), 577.

Lúquez, C., Raphael, B. H., Joseph, L. A., Meno, S. R., Fernández, R. A., & Maslanka, S. E. (2012). Genetic diversity among *Clostridium botulinum* strains harboring bont/A2 and bont/A3 genes. *Applied and Environmental Microbiology*, *78*(24), 8712–8718.

MacDonald, E., Arnesen, T. M., Brantsaeter, A. B., Gerlyng, P., Grepp, M., Hansen, B. Å., ... & Vold, L. (2013). Outbreak of wound botulism in people who inject drugs, Norway, October to November 2013. *Euro surveillance: bulletin*

Européen sur les maladies transmissibles= European communicable disease bulletin, *18*(45).

MacDonald, T. E., Helma, C. H., Ticknor, L. O., Jackson, P. J., Okinaka, R. T., Smith, L. A., … Hill, K. K. (2008). Differentiation of *Clostridium botulinum* serotype A strains by multiple-locus variable-number tandem-repeat analysis. *Applied and Environmental Microbiology*, *74*(3), 875–882.

Mackey, B. M. (2000). Injured bacteria. In: Lund, B., Baird-Parker, T. C., & Gould, G. W. (Eds.). *Microbiological Safety and Quality of Food* (Vol. 1). Springer. pp., 315-341.

Magoc, T., Pabinger, S., Canzar, S., Liu, X., Su, Q., Puiu, D., ... & Salzberg, S. L. (2013). GAGE-B: an evaluation of genome assemblers for bacterial organisms. *Bioinformatics*, *29*(14), 1718-1725.

Mah, J. H., Kang, D. H., & Tang, J. (2008a). Morphological study of heat-sensitive and heat-resistant spores of *Clostridium sporogenes*, using transmission electron microscopy. *Journal of Food Protection*, *71*(5), 953-958.

Mah, J. H., Kang, D. H., & Tang, J. (2008b). Effects of minerals on sporulation and heat resistance of *Clostridium sporogenes*. *International Journal of Food Microbiology*, *128*(2), 385-389.

Marie, D., Vaulot, D., & Partensky, F. (1996). Application of the novel nucleic acid dyes YOYO-1, YO-PRO-1, and PicoGreen for flow cytometric analysis of marine prokaryotes. *Applied and Environmental Microbiology*, *62*(5), 1649-1655.

Markowitz, V. M., Chen, I. M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., ... & Kyrpides, N. C. (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research*, *40*(D1), D115-D122.

Marquis, R. E., & Shin, S. Y. (1994). Mineralization and responses of bacterial spores to heat and oxidative agents. *FEMS Microbiology Reviews*, *14*(4), 375–379.

Marshall, K. M., Bradshaw, M., & Johnson, E. A. (2010). Conjugative botulinum neurotoxin-encoding plasmids in *Clostridium botulinum*. *PLoS One*, *5*(6), e11087.

Marvaud, J. C., Gibert, M., Inoue, K., Fujinaga, Y., Oguma, K., & Popoff, M. R. (1998). botR/A is a positive regulator of botulinum neurotoxin and associated non-toxin protein genes in *Clostridium botulinum* A. *Molecular Microbiology*, *29*(4), 1009-1018.

Mathys, A., Chapman, B., Bull, M., Heinz, V., & Knorr, D. (2007). Flow cytometric assessment of *Bacillus* spore response to high pressure and heat. *Innovative Food Science & Emerging Technologies*, *8*(4), 519-527.

Matsumura, T., Jin, Y., Kabumoto, Y., Takegahara, Y., Oguma, K., Lencer, W. I., & Fujinaga, Y. (2008). The HA proteins of botulinum toxin disrupt intestinal epithelial intercellular junctions to increase toxin absorption. *Cellular Microbiology*, *10*(2), 355–364.

Matz, L. L., Beaman, T. C., & Gerhardt, P. (1970). Chemical composition of exosporium from spores of *Bacillus cereus*. *Journal of Bacteriology*, *101*(1), 196–201.

McCroskey, L. M., Hatheway, C. L., Fenicia, L., Pasolini, B., & Aureli, P. (1986). Characterization of an organism that produces type E botulinal toxin but which resembles *Clostridium butyricum* from the feces of an infant with type E botulism. *Journal of Clinical Microbiology*, *23*(1), 201–202.

McLauchlin, J., Grant, K. A., & Little, C. L. (2006). Food-borne botulism in the United Kingdom. *Journal of Public Health*, *28*(4), 337-342.

Medini, D., Donati, C., Tettelin, H., Masignani, V., & Rappuoli, R. (2005). The microbial pan-genome. *Current Opinion in Genetics & Development*, *15*(6), 589-594.

van Melis, C. C. J., Nierop Groot, M. N., Tempelaars, M. H., Moezelaar, R., & Abee, T. (2011). Characterization of germination and outgrowth of sorbic acid-stressed *Bacillus cereus* ATCC 14579 spores: Phenotype and transcriptome analysis. *Food Microbiology*, *28*(2), 275–283.

Melly, E., Genest, P. C., Gilmore, M. E., Little, S., Popham, D. L., Driks, A., & Setlow, P. (2002). Analysis of the properties of spores of *Bacillus subtilis* prepared at different temperatures. *Journal of Applied Microbiology*, *92*(6), 1105-1115.

Metchnikoff, E. (1908). Études sur la flore intestinale. *Annales de l'Institut Pasteur*, *22*, 929–955.

Mills, G. Earnshaw, R. & Patterson, M.F. (1998). Effects of high hydrostatic pressure on *Clostridium sporogenes* spores. *Letters in Applied Microbiology*, *26*(3), 227–230.

Mira, A., Martín-Cuadrado, A. B., D'Auria, G., & Rodríguez-Valera, F. (2010). The bacterial pan-genome: a new paradigm in microbiology. *International Microbiology*, *13*(2), 45-57.

Moir, C.J., Murrell, W.G., Richardson, K.C. & Board, P.W. (2001). Commercially sterile foods. In: Moir, C.J., Andrew-Kabilafkas, C., Arnold, G., Cox, B.M., Hocking, A.D. and Jenson, I., (Eds.) *Spoilage of processed foods : causes and diagnosis*, pp 101-112. Australia. AIFST. NSW Branch Food Microbiology Group.

Moir, A. (2003). Bacterial spore germination and protein mobility. *Trends in Microbiology*, *11*(10), 452–454.

Moir, A. (2006). How do spores germinate? *Journal of Applied Microbiology*, *101*(3), 526–530.

Molle, V., Fujita, M., Jensen, S. T., Eichenberger, P., González-Pastor, J. E., Liu, J. S., & Losick, R. (2003). The Spo0A regulon of *Bacillus subtilis*. *Molecular Microbiology*, *50*(5), 1683-1701.

Montville, T. J. (1984). Quantitation of pH- and salt-tolerant subpopulations from *Clostridium botulinum*. *Applied and Environmental Microbiology*, *47*(1), 28 –30.

NACMCF: National Advisory Committee on Microbiological Criteria for Foods. (2010). Parameters for determining inoculated pack/challenge study protocols. *Journal of Food Protection*, *73*(1), 140–202.

Nakashio, S., & Gerhardt, P. (1985). Protoplast dehydration correlated with heat resistance of bacterial spores. *Journal of Bacteriology*, *162*(2), 571-578.

Nale, J. Y., Shan, J., Hickenbotham, P. T., Fawley, W. N., Wilcox, M. H., & Clokie, M. R. (2012). Diverse temperate bacteriophage carriage in *Clostridium difficile* 027 strains. *PloS one*, *7*(5), e37263.

Nevas, M., Lindstrom, M., Hielm, S., Bjorkroth, K. J., Peck, M. W., & Korkeala, H. (2005). Diversity of proteolytic *Clostridium botulinum* strains, determined by a pulsed-field gel electrophoresis approach. *Applied and Environmental Microbiology*, *71*(3), 1311–1317.

Nicholson, W. L., Munakata, N., Horneck, G., Melosh, H. J., & Setlow, P. (2000). Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiology and Molecular Biology Reviews*, *64*(3), 548–572.

Nicholson, W. L., Setlow, P. (1990). Sporulation, germination and outgrowth. In: Harwood C R, Cutting S M (Eds.), *Molecular Microbiological Methods for Bacillus*. Chichester, United Kingdom: John Wiley & Sons, Ltd. pp. 391–450.

Olsen, A. M., & Scott, W. J. (1946). Influence of starch in media used for the detection of heated bacterial spores. *Nature*, *157*(3985), 337–337.

Olsen, J. S., Scholz, H., Fillo, S., Ramisse, V., Lista, F., Trømborg, A. K., … Blatny, J. M. (2014). Analysis of the genetic distribution among members of *Clostridium botulinum* group I using a novel multilocus sequence typing (MLST) assay. *Journal of Microbiological Methods*, *96*, 84–91.

Paidhungat, M., & Setlow, P. (2001). Localization of a germinant receptor protein (GerBA) to the inner membrane of *Bacillus subtilis* spores. *Journal of Bacteriology*, *183*(13), 3982–3990.

Peck, M. W. (2009). Biology and genomic analysis of *Clostridium botulinum*. In: Robert K. Poole (Ed.), *Advances in Microbial Physiology* (Vol. Volume 55, pp. 183–320). Academic Press.

Peck, M. W. (2006). *Clostridium botulinum* and the safety of minimally heated, chilled foods: an emerging issue? *Journal of Applied Microbiology*, *101*(3), 556–570.

Peck, M. w., Fairbairn, D. a., & Lund, B. M. (1992). Factors affecting growth from heat-treated spores of non-proteolytic *Clostridium botulinum*. *Letters in Applied Microbiology*, *15*(4), 152–155.

Peleg, M. (1996). Evaluation of the Fermi equation as a model of dose-response curves. *Applied Microbiology and Biotechnology*, *46*(3), 303–306.

Petkau, A., Stuart-Edwards, M., Stothard, P., & Van Domselaar, G. (2010). Interactive microbial genome visualization with GView. *Bioinformatics*, *26*(24), 3125-3126.

Pflug, I. J. (2010). Science, practice, and human errors in controlling *Clostridium botulinum* in heat-preserved food in hermetic containers. *Journal of Food Protection*, *73*(5), 993–1002.

Popham, D. L., Helin, J., Costello, C. E., & Setlow, P. (1996). Analysis of the peptidoglycan structure of *Bacillus subtilis* endospores. *Journal of Bacteriology*, *178*(22), 6451–6458.

Popham, D.L., & Setlow, P. (1993) Popham, D. L., Illades-Aguiar, B., & Setlow, P. (1995). The *Bacillus subtilis dacB* gene, encoding penicillin-binding protein 5*, is part of a three-gene operon required for proper spore cortex synthesis and spore core dehydration. *Journal of Bacteriology*, *177*(16), 4721-4729.

Poxton, I. R. (1984). Demonstration of the common antigens *of Clostridium botulinum, C. sporogenes* and *C. novyi* by an enzyme-linked immunosorbent assay and electroblot transfer. *Journal of General Microbiology*, *130*(4), 975-981.

Pourshafie, M., Vahdani, P., & Popoff, M. (2005). Genotyping *Clostridium botulinum* toxinotype A isolates from patients using amplified rDNA restriction analysis. *Journal of Medical Microbiology*, *54*(10), 933–936.

Powell, J. F., & Hunter, J. R. (1955). Spore germination in the genus *Bacillus*: the modification of fermination eequirements as a eesult of preheating. *Journal of General Microbiology*, *13*(1), 59–67.

Prévot, A. R. (1953). Rapport d'introduction du President du souscomité *Clostridium* pour l'unification de la nomenclature des types toxinogènes de *C. botulinum*. *International Bulletin of Bacteriological Nomenclature and Taxonomy*, *3*(2-3), 120–123.

Raffestin, S., Dupuy, B., Marvaud, J. C., & Popoff, M. R. (2005). BotR/A and TetR are alternative RNA polymerase sigma factors controlling the expression of the neurotoxin and associated protein genes in *Clostridium botulinum* type A and *Clostridium tetani*. *Molecular Microbiology, 55*(1), 235–249.

Ragkousi, K., Cowan, A. E., Ross, M. A., & Setlow, P. (2000). Analysis of nucleoid morphology during germination and outgrowth of spores of *Bacillus* species. *Journal of Bacteriology, 182*(19), 5556-5562.

Ramaswamy, H. S., & Shao, Y. (2010). High pressure destruction kinetics of *Clostridium sporogenes* spores in salmon slurry at elevated temperatures.*International Journal of Food Properties*, *13*(5), 1074-1091.

Raphael, B. H., Joseph, L. A., McCroskey, L. M., Lúquez, C., & Maslanka, S. E. (2010). Detection and differentiation of *Clostridium botulinum* type A strains using a focused DNA microarray. *Molecular and Cellular Probes, 24*(3), 146–153.

Rasko, D. A., Rosovitz, M. J., Myers, G. S., Mongodin, E. F., Fricke, W. F., Gajer, P., ... & Ravel, J. (2008). The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *Journal of Bacteriology*, *190*(20), 6881-6893.

Reddy, N. R., Tetzloff, R. C., & Skinner, G. E. (2010). Effect of media, additives, and incubation conditions on the recovery of high pressure and heat-injured *Clostridium botulinum* spores. *Food microbiology*, *27*(5), 613-617.

Riemann, H., & Ordal, Z. J. (1961). Germination of bacterial endospores with calcium and dipicolinic acid. *Science, 133*(3465), 1703–1704.

Roberts, T. A., Gilbert, R. J., & Ingram, M. (1966). The effect of sodium chloride on heat resistance and recovery of heated spores of *Clostridium sporogenes* (PA 3679/S2). *Journal of Applied Microbiology*, *29*(3), 549-555.

Rode, L. J., Pope, L., Filip, C., & Smith, L. D. (1971). Spore appendages and taxonomy of *Clostridium sordellii*. *Journal of Bacteriology, 108*(3), 1384–1389.

Roberts, C. M., & Hoover, D. G. (1996). Sensitivity of *Bacillus coagulans* spores to combinations of high hydrostatic pressure, heat, acidity and nisin. *Journal of Applied Microbiology*, *81*(4), 363-368.

Rodriguez, A. C., Smerage, G. H., Teixeira, A. A., Lindsay, J. A., & Busta, F. F. (1992). Population model of bacterial spores for validation of dynamic thermal processes. *Journal of Food Process Engineering*, *15*(1), 1-30.

Ross, C., & Abel-Santos, E. (2010). The Ger receptor family from sporulating bacteria. *Current Issues in Molecular Biology*, *12*(3), 147.

Russell, A. D. (1990). Bacterial spores and chemical sporicidal agents. *Clinical Microbiology Reviews*, *3*(2), 99–119.

Sakaguchi, Y., Hayashi, T., Kurokawa, K., Nakayama, K., Oshima, K., Fujinaga, Y., ... & Oguma, K. (2005). The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(48), 17472-17477.

Santo, L. Y., & Doi, R. H. (1974). Ultrastructural Analysis During Germination and Outgrowth of *Bacillus subtilis* Spores. *Journal of Bacteriology*, *120*(1), 475–481.

Sapru, V., Smerage, G. H., Teixeira, A. A., & Lindsay, J. A. (1993). Comparison of predictive models for bacterial spore population resources to sterilization temperatures. *Journal of Food Science*, *58*(1), 223-228.

Savani, J., Harris, N. D., & Gould, W. A. (1978). Survival of *Clostridium sporogenes* PA 3679 in home-canned tomatoes. *Journal of Food Science*, *43*(1), 222-224.

Schuch, R., & Fischetti, V. A. (2009). The secret life of the anthrax agent *Bacillus anthracis*: bacteriophage-mediated ecological adaptations. *PLoS One*, *4*(8), e6532.

Scott, V. N., & Bernard, D. T. (1985). The Effect of Lysozyme on the apparent heat resistance of nonproteolytic type B *Clostridium Botulinum*. *Journal of Food Safety*, *7*(3), 145–154.

Sebaihia, M., Peck, M. W., Minton, N. P., Thomson, N. R., Holden, M. T. G., Mitchell, W. J. & Parkhill, J. (2007). Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. *Genome Research*, *17*(7), 1082–1092.

Setlow, P. (2007). I will survive: DNA protection in bacterial spores. *Trends in Microbiology*, *15(4)*, 172-180.

Setlow, P. (2006). Spores of *Bacillus subtilis*: their resistance to and killing by radiation, heat and chemicals. *Journal of Applied Microbiology*, *101*(3), 514–525.

Setlow, P. (2003). Spore germination. *Current Opinion in Microbiology*, *6*(6), 550–556.

Setlow, B., Loshon, C. A., Genest, P. C., Cowan, A. E., Setlow, C., & Setlow, P. (2002). Mechanisms of killing spores of *Bacillus subtilis* by acid, alkali and ethanol. *Journal of Applied Microbiology*, *92*(2), 362-375.

Setlow, P. (2000). Resistance of bacterial spores. In: Storz, G. & Hengge-Aronis, R. (Eds.). *Bacterial Stress Responses*, ASM Press, Washington, DC, pp. 217-230.

Setlow, P. & Johnson, E. A. (1997). Spores and their significance. In: M.P. Doyle, L.R. Beuchat & T.J. Montville, (Eds.), *Food Microbiology Fundamentals and Frontiers*, ASM Press, Washington, DC, pp. 30–65.

Shah, I. M., Laaberki, M. H., Popham, D. L., & Dworkin, J. (2008). A eukaryotic-like *Ser/Thr* kinase signals bacteria to exit dormancy in response to peptidoglycan fragments. *Cell*, *135*(3), 486-496.

Shi, G. M., Xu, Y., Fan, J., Zhou, J., Yang, X. R., Qiu, S. J., & Wu, Z. Q. (2008). Identification of side population cells in human hepatocellular carcinoma cell lines with stepwise metastatic potentials. *Journal of Cancer Research and Clinical Oncology*, *134*(11), 1155-1163.

Shull, J. J., Cargo, G. T., & Ernst, R. R. (1963). Kinetics of heat activation and of thermal death of bacterial spores. *Applied Microbiology*, *11*(6), 485-487.

Skarin, H., Håfström, T., Westerberg, J., & Segerman, B. (2011). *Clostridium botulinum* group III: a group with dual identity shaped by plasmids, phages and mobile elements. *BMC Genomics*, *12*(1), 185.

Skinner, G. E., Gendel, S. M., Fingerhut, G. A., Solomon, H. A., & Ulaszek, J. (2000). Differentiation between types and strains of *Clostridium botulinum* by riboprinting. *Journal of Food Protection*, *63*(10), 1347–1352.

Smelt, J. P. P. M., Bos, A. P., Kort, R., & Brul, S. (2008). Modelling the effect of sub(lethal) heat treatment of *Bacillus subtilis* spores on germination rate and outgrowth to exponentially growing vegetative cells. *International Journal of Food Microbiology*, *128*(1), 34–40.

Smith, T. J., Hill, K. K., Foley, B. T., Detter, J. C., Munk, A. C., Bruce, D. C., … Brettin, T. S. (2007). Analysis of the Neurotoxin Complex Genes in *Clostridium botulinum* A1-A4 and B1 Strains: BoNT/A3, /Ba4 and /B1 Clusters Are Located within Plasmids. *PLoS ONE*, *2*(12), e1271.

Smith, T. J., Lou, J., Geren, I. N., Forsyth, C. M., Tsai, R., LaPorte, S. L., … Marks, J. D. (2005). Sequence Variation within Botulinum Neurotoxin Serotypes Impacts Antibody Binding and Neutralization. *Infection and Immunity*, *73*(9), 5450–5457.

Sneath, P. H. (1986). Endospore-forming Gram-positive rods and cocci. *Bergey 0s Manual of Systematic Bacteriology, Sneath, PHA et al.,(ed.)(Baltimore: Williams & Wilkins)*, 1104–1207.

Stackebrandt, E., & Goebel, B. M. (1994). Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology*, *44*(4), 846–849.

Stevenson, K. E., & Vaughn, R. H. (1972). Exosporium formation in sporulating cells of *Clostridium botulinum* 78A. *Journal of bacteriology*, *112*(1), 618-621.

Stocks, S. M. (2004). Mechanism and use of the commercially available viability stain, BacLight. *Cytometry Part A*, *61*(2), 189-195.

Stone, G., Chapman, B., & Lovell, D. (2009). Development of a log-quadratic model to describe microbial inactivation, illustrated by thermal inactivation of *Clostridium botulinum*. *Applied and Environmental Microbiology*, *75*(22), 6998–7005.

Stragier, P. (2002). Ch.35. A Gene Odyssey: Exploring the genomes of endospore-forming bacteria. In: Sonenshein, A.L., Hoch, J.A. & Losick, R. (Eds.). *Bacillus subtilis and its closest relatives*, ASM Press.

Sträuber, H., & Müller, S. (2010). Viability states of bacteria—specific mechanisms of selected probes. *Cytometry Part A*, *77*(7), 623-634.

Stringer, S. C., Webb, M. D., George, S. M., Pin, C., & Peck, M. W. (2005). Heterogeneity of times required for germination and outgrowth from single spores of nonproteolytic *Clostridium botulinum*. *Applied and Environmental Microbiology*, *71*(9), 4998–5003.

Stringer, S. C., Webb, M. D., & Peck, M. W. (2011). Lag time variability in individual spores of *Clostridium botulinum*. *Food Microbiology*, *28*(2), 228–235.

Stumbo, C. R. (1973). *Thermobacteriology in Food Processing*. Academic Press.

Suen, J. C., Hatheway, C. L., Steigerwalt, A. G., & Brenner, D. J. (1988). *Clostridium argentinense* sp. nov.: A genetically homogeneous group composed of all strains of *Clostridium botulinum* toxin type G and some nontoxigenic strains previously identified as *Clostridium subterminale* or *Clostridium hastiforme*. *International Journal of Systematic Bacteriology*, *38*(4), 375–381.

Swain, M. T., Tsai, I. J., Assefa, S. A., Newbold, C., Berriman, M., & Otto, T. D. (2012). A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nature Protocols*, *7*(7), 1260-1284.

Takubo, Y., Atarashi, M., Nishihara, T., & Kondo, M. (1988). Isolation and characterization of outermost layer deficient mutant spores of *Bacillus megaterium*. *Microbiology and immunology*, *32*(9), 973-979.

Tamborrini, M., Oberli, M. a., Werz, D. b., Schürch, N., Frey, J., Seeberger, P. h., & Pluschke, G. (2009). Immuno-detection of anthrose containing tetrasaccharide in the exosporium of *Bacillus anthracis* and *Bacillus cereus* strains. *Journal of Applied Microbiology*, *106*(5), 1618–1628.

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*, *10*(3), 512-526.

Taylor, R. H., Dunn, M. L., Ogden, L. V., Jefferies, L. K., Eggett, D. L., & Steele, F. M. (2013). Conditions associated with *Clostridium sporogenes* growth as a

surrogate for *Clostridium botulinum* in nonthermally processed canned butter. *Journal of Dairy Science*, *96*(5), 2754–2764.

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., ... & Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 13950-13955.

Tipper, D. J., & Gauthier, J. J. (1972). Structure of the bacterial endospore. *Spores V. American Society for Microbiology, Washington, DC*, 3–12.

Townsend, C. T., Esty, J. K., & Baselt, F. C. (1938). Heat-resistance studies on spores of putrefactive anaerobes in relation to determination of safe processes for canned foods. *Journal of Food Science*, *3*(3), 323–346.

Tracy, B. P., Gaida, S. M., & Papoutsakis, E. T. (2008). Development and application of flow-cytometric techniques for analyzing and sorting endospore-forming clostridia. *Applied and environmental microbiology*, *74*(24), 7497-7506.

Umeda, K., Wada, T., Kohda, T., & Kozaki, S. (2013). Multi-locus variable number tandem repeat analysis for *Clostridium botulinum* type B isolates in Japan: comparison with other isolates and genotyping methods. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *16*, 298–304.

Umeda, K., Seto, Y., Kohda, T., Mukamoto, M., & Kozaki, S. (2012). Stability of toxigenicity in proteolytic *Clostridium botulinum* type B upon serial passage. *Microbiology and Immunology*, *56*(5), 338–341.

Umeda, K., Seto, Y., Kohda, T., Mukamoto, M., & Kozaki, S. (2009). Genetic characterization of *Clostridium botulinum* associated with type B infant botulism in Japan. *Journal of Clinical Microbiology*, *47*(9), 2720-2728.

Vanhomwegen, J., Berthet, N., Mazuet, C., Guigon, G., Vallaeys, T., Stamboliyska, R., … Popoff, M.-R. (2013). Application of high-density DNA resequencing microarray for detection and characterization of botulinum neurotoxin-producing Clostridia. *PLoS ONE*, *8*(6), e67510.

Veal, D. A., Deere, D., Ferrari, B., Piper, J., & Attfield, P. V. (2000). Fluorescence staining and flow cytometry for monitoring microbial cells. *Journal of Immunological Methods*, *243*(1), 191-210.

Veening, J.-W., Smits, W. K., & Kuipers, O. P. (2008). Bistability, epigenetics, and bet-hedging in bacteria. *Annual Review of Microbiology*, *62*(1), 193–210.

de Vries, 2004. *Bacillus cereus* spore formation, structure and germination. Thesis. Wageningen University and Research Centre, Wageningen, The Netherlands.

Waites, W. M., Stansfield, R., & Bayliss, C. E. (1979). The effect of sporulation medium on the structure and heat resistance of spores of *Closridium bifermentans*. *FEMS Microbiology Letters*, *5*(5), 365-368.

Walker, J. R., Gnanam, A. J., Blinkova, A. L., Hermandson, M. J., Karymov, M. A., Lyubchenko, Y. L., … Linse, K. D. (2007). *Clostridium taeniosporum* spore ribbon-like appendage structure, composition and genes. *Molecular Microbiology*, *63*(3), 629–643.

Webb, M. D., Pin, C., Peck, M. W., & Stringer, S. C. (2007). Historical and contemporary NaCl concentrations affect the duration and distribution of lag times from individual spores of nonproteolytic *Clostridium botulinum*. *Applied and environmental microbiology*, *73*(7), 2118-2127.

Wiencek, K. M., Klapes, N. A., & Foegeding, P. M. (1990). Hydrophobicity of *Bacillus* and *Clostridium* spores. *Applied and Environmental Microbiology*, *56*(9), 2600-2605.

Williams, O. B., & Reed, J. M. (1942). The significance of the incubation temperature of recovery cultures in determining spore resistance to heat. *Journal of Infectious Diseases*, *71*(3), 225-227.

Wörner, K., Szurmant, H., Chiang, C., & Hoch, J. A. (2006). Phosphorylation and functional analysis of the sporulation initiation factor Spo0A from *Clostridium botulinum*. *Molecular Microbiology*, *59*(3), 1000-1012.

Xiao, Y., Francke, C., Abee, T., & Wells-Bennik, M. H. J. (2011). Clostridial spore germination versus bacilli: genome mining and current insights. *Food Microbiology*, *28*(2), 266–274.

Yuan, J., Inami, G., Mohle-Boetani, J., & Vugia, D. J. (2011). Recurrent wound botulism among injection drug users in California. *Clinical Infectious Diseases*, *52*(7), 862-866.

Zeigler, D. R. (2003). Gene sequences useful for predicting relatedness of whole genomes in bacteria. *International Journal of Systematic and Evolutionary Microbiology*, *53*(6), 1893-1900.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821-829.

Zhang, L., Ma, H., & Pugh, B. F. (2011). Stable and dynamic nucleosome states during a meiotic developmental process. *Genome Research*, *21*(6), 875-884.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Research*, *39*(suppl 2), W347-W352.

Zhou, Y., Sugiyama, H., Nakano, H., & Johnson, E. A. (1995). The genes for the *Clostridium botulinum* type G toxin complex are on a plasmid. *Infection and Immunity*, *63*(5), 2087–2091.

Appendix A. Custom bash scripts (*Clean_reads.sh* & *Align.sh*) used for the quality control of MiSeq and HiSeq reads and estimation of insert size prior to *de novo* assembly.

#Clean_reads.sh

```
#!/bin/bash
#This script is designed to take input of paired end MiSeq or HiSeq reads in either fastq or fastq.gz format and output an interleaved fastq file
#Post quality trimming using Trimmoatic this script prepares a Velvet output interleaved fastq file with adjusted header values for forward and reverse strand reads
#This script is written for MiSeq reads, using for other reads may require adjustment of the custom trimming settings and require a custom adapter fasta file.

#CUSTOM TRIMMING SETTINGS see http://www.usadellab.org/cms/index.php?page=trimmomatic for explanations of settings
TRIMPATH=~/Bioinformatics/Trimmomatic-0.30/trimmomatic-0.30.jar
LEADING=3
TRAILING=3
HEADCROP=20
CROP=220
SLIDINGWINDOW=4:20
MINLEN=180

if [ ! -f $1_L001_R1_001.fastq.gz ] && [ ! -f $1_L001_R1_001.fastq ] ; then
   echo "$1_L001_R1_001.fastq(.gz) - File not found!"
   exit 0
fi
if [ ! -f $TRIMPATH ] ; then
   echo "Trimmoatic (http://www.usadellab.org/cms/index.php?page=trimmomatic) not installed"
   exit 0
fi

#CUSTOM ADAPTER SEQUENCES - ADD file "adapters.fa" to working directory for custom adapters
if [ ! -f adapters.fa ]; then
echo ">Adapter_index_A
CTGTCTCTTATACACATCTCCGAGCCCACGAGAC
> Adapter_index_A
CTGTCTCTTATACACATCTGACGCTGCCGACGA" > adapters_temp.fa
else
cp adapters.fa adapters_temp.fa;
fi

#Detecting fasta or fasta.gz
if [ ! -f $1_L001_R1_001.fastq.gz ]; then
    EXT=fastq
else
    EXT=fastq.gz
fi

#Running trimmomatic
java -jar $TRIMPATH PE -phred33 $1_L001_R1_001.$EXT $1_L001_R2_001.$EXT $1_R1_temp.fastq $1_R1_unpaired_temp.fastq $1_R2_temp.fastq $1_R2_unpaired_temp.fastq ILLUMINACLIP:adapters_temp.fa:3:7:7 LEADING:$LEADING TRAILING:$TRAILING HEADCROP:${3:-$HEADCROP} CROP:${4:-$CROP} SLIDINGWINDOW:4:20 MINLEN:${2:-$MINLEN}

#Addition of header "/1" or "/2"information for velvet
sed '1~4 s:$:\/1:g' $1_R1_temp.fastq>$1_R1.fastq
sed '1~4 s:$:\/1:g' $1_R1_unpaired_temp.fastq>$1_R1_unpaired.fastq
sed '1~4 s:$:\/2:g' $1_R2_temp.fastq>$1_R2.fastq
sed '1~4 s:$:\/2:g' $1_R2_unpaired_temp.fastq>$1_R2_unpaired.fastq

#Interleave paired data (courtesy:https://gist.github.com/4544979)
```

```
paste $1_R1.fastq $1_R2.fastq | paste - - - - | awk -v OFS="\n" -v FS="\t" '{print($1,$3,$5,$7,$2,$4,$6,$8)}'> $1_shuffled.fastq

#Cleanup of temporary files
rm adapters_temp.fa
rm $1_R1_temp.fastq
rm $1_R1_unpaired_temp.fastq
rm $1_R2_temp.fastq
rm $1_R2_unpaired_temp.fastq
```

#Align.sh

```
#!/bin/bash
#This script aligns pre-processed reads using bwa mem before using Picard to estimate the insert size for downstream assembly
#Alignment of reads using bwa mem
bwa index $2.fna
bwa mem -p $2.fna $1_shuffled.fastq > $1_$2.sam
samtools view -bS $1_$2.sam > a.bam
samtools sort a.bam $1_$2

#Cleanup of temporary files
rm $2.fna.amb
rm $2.fna.ann
rm $2.fna.bwt
rm $2.fna.pac
rm $2.fna.sa
rm a.bam
rm $1_$2.sam

#Determination of insert size using Picard tools
IS=~/Bioinformatics/picard-tools-1.101/CollectInsertSizeMetrics.jar
if [ ! -f $1_$2.bam ] ; then
    echo "$1_$2.bam - File not found!"
    exit 0
fi
if [ ! -f $IS ] ; then
    echo "Picard (or path) not correctly installed"
    exit 0
fi

#Running Picard
java -jar $IS H=histogram_$1_$2.pdf I=$1_$2.bam O=Insertsize_$1_$2.txt
```

Appendix B. Table of genes (n=4111) identified in the genome of *Clostridium sporogenes* PA 3679 (NCTC 8594) using the NCBI prokaryotic genome annotation pipeline and IMG pipeline.

locus_tag	Minimum	Maximum	Length	Product	COG Group
Genomic Scaffold: NZ_JH470455					
IYC_00005	64	201	138	hypothetical protein	
IYC_00010	348	512	165	hypothetical protein	
IYC_00015	527	964	438	Gp50 protein	COG1403 Restriction endonuclease
IYC_00020	979	1,287	309	hypothetical protein	
IYC_00025	1,302	1,571	270	hypothetical protein	
IYC_00030	1,687	1,920	234	hypothetical protein	
IYC_00035	1,886	2,257	372	hypothetical protein	
IYC_00040	2,432	2,557	126	hypothetical protein	COG1403 Restriction endonuclease
IYC_00045	2,704	3,090	387	phage terminase	COG3747 Phage terminase, small subunit
IYC_00065	4,789	4,962	174	hypothetical protein	
IYC_00070	4,967	6,193	1,227	phage portal protein, HK97 family	COG4695 Phage-related protein
IYC_00075	6,193	6,876	684	phage protein	COG0740 Protease subunit of ATP-dependent Clp proteases
IYC_00080	6,920	8,065	1,146	phage protein	COG4653 Predicted phage phi-C31 gp36 major capsid-like protein
IYC_00085	8,076	8,219	144	hypothetical protein	
IYC_00090	8,219	8,518	300	hypothetical protein	
IYC_00095	8,598	8,933	336	hypothetical protein	
IYC_00100	8,943	9,152	210	major surface protein 3	
IYC_00105	9,154	9,564	411	hypothetical protein	
IYC_00110	9,577	10,035	459	hypothetical protein	
IYC_00115	10,217	11,125	909	hypothetical protein	
IYC_00120	11,137	11,559	423	hypothetical protein	
IYC_00125	11,570	11,983	414	hypothetical protein	
IYC_00130	12,234	14,765	2,532	phage tail tape measure protein, TP901 family, core region	COG5283 Phage-related tail protein
IYC_00135	14,811	15,242	432	hypothetical protein	
IYC_00140	15,242	16,171	930	hypothetical protein	
IYC_00145	16,184	16,501	318	hypothetical protein	
IYC_00150	16,513	16,950	438	hypothetical protein	
IYC_00155	16,962	18,071	1,110	hypothetical protein	COG3299 Uncharacterized homolog of phage Mu proteingp47
IYC_00160	18,084	18,566	483	hypothetical protein	
IYC_00165	18,622	19,059	438	hypothetical protein	
IYC_00170	19,072	19,296	225	hypothetical protein	
Genomic Scaffold: NZ_JH470456					
IYC_00175	<1	137	>137	hypothetical protein	
IYC_00180	151	1,320	1,170	phage integrase family site specific recombinase	COG0582 Integrase
IYC_00185	1,375	1,707	333	DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_00190	1,909	2,154	246	DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_00195	2,126	2,401	276	excisionase family DNA-binding protein	
IYC_00200	2,415	>2664	>250	hypothetical protein	COG3617 Prophage antirepressor
Genomic Scaffold: NZ_JH470457					
IYC_00205	<1	231	>231	phage protein	
IYC_00210	228	389	162	hypothetical protein	
IYC_00215	402	770	369	phage protein	
IYC_00220	890	1,483	594	Phage protein	
IYC_00225	1,486	1,719	234	hypothetical protein	
IYC_00230	1,731	2,012	282	hypothetical protein	COG3956 Protein containing tetrapyrrole methyltransferase domain and MazG-like (predicted pyrophosphatase) domain
IYC_00235	2,012	2,251	240	SpoVT/AbrB family regulatory protein	COG2002 Regulators of stationary/sporulation gene expression
IYC_00240	2,283	3,056	774	hypothetical protein	COG1484 DNA replication protein
IYC_00245	3,046	3,837	792	hypothetical protein	COG3935 Putative primosome component and related proteins
IYC_00250	3,859	4,143	285	hypothetical protein	
IYC_00255	4,145	4,741	597	hypothetical protein	
IYC_00260	4,741	5,232	492	hypothetical protein	
IYC_00265	5,216	5,362	147	hypothetical protein	
IYC_00270	5,362	>5483	>122	hypothetical protein	
Genomic Scaffold: NZ_JH470458					
IYC_00275	100	240	141	hypothetical protein	
IYC_00280	218	547	330	hypothetical protein	COG1363 Cellulase M and related proteins
IYC_00285	664	864	201	hypothetical protein	
IYC_00290	1,315	1,596	282	hypothetical protein	
IYC_00295	1,597	1,893	297	hypothetical protein	COG1503 Peptide chain release factor 1 (eRF1)
IYC_00300	1,916	2,299	384	hypothetical protein	
IYC_00305	2,303	3,160	858	hypothetical protein	
IYC_00310	3,161	3,382	222	hypothetical protein	
IYC_00315	3,469	3,648	180	hypothetical protein	
IYC_00320	3,658	4,191	534	hypothetical protein	
IYC_00325	4,192	4,407	216	hypothetical protein	
IYC_00330	4,391	4,555	165	hypothetical protein	
IYC_00335	4,571	4,711	141	hypothetical protein	
IYC_00340	4,713	4,850	138	hypothetical protein	
IYC_00345	4,882	5,190	309	hypothetical protein	
IYC_00350	5,190	>5340	>151	phage protein	COG4570 Holliday junction resolvase
Genomic Scaffold: NZ_JH470459					
IYC_00357	29	493	465	AsnC family transcriptional regulator	COG1522 Transcriptional regulators
IYC_00362	1,032	1,391	360	hypothetical protein	
IYC_00367	2,022	3,395	1,374	amino acid permease family protein	COG0531 Amino acid transporters
IYC_00372	3,888	5,237	1,350	amino acid permease family protein	COG0531 Amino acid transporters
IYC_00377	5,579	6,088	510	hexapeptide repeat-containing transferase	COG0663 Carbonic anhydrases/acetyltransferases, isoleucine patch superfamily
IYC_00382	6,133	7,005	873	metallo-beta-lactamase family protein	COG2333 Predicted hydrolase (metallo-beta-lactamasesuperfamily)
IYC_00387	7,051	9,669	2,619	calcium-translocating P-type ATPase, PMCA-type	COG0474 Cation transport ATPase
IYC_00392	10,042	10,578	537	chromate transporter protein	COG2059 Chromate transport protein ChrA
IYC_00397	10,578	11,114	537	chromate transport protein	COG2059 Chromate transport protein ChrA
IYC_00402	11,288	11,878	591	PemK family protein	COG2337 Growth inhibitor
IYC_00407	12,319	12,744	426	hypothetical protein	COG1728 Uncharacterized protein conserved in bacteria
IYC_00412	12,951	14,873	1,923	NAD synthetase	COG0388 Predicted amidohydrolase
IYC_00417	15,098	15,538	441	hypothetical protein	COG1585 Membrane protein implicated in regulation of membrane protease activity
IYC_00422	15,541	16,479	939	spfh domain/band 7 family protein	COG0330 Membrane protease subunits, stomatin/prohibitin homologs
IYC_00427	16,973	18,208	1,236	xaa-pro aminopeptidase	COG0006 Xaa-Pro aminopeptidase
Genomic Scaffold: NZ_JH470460					
IYC_00432	<1	475	>475	hypothetical protein	
IYC_00437	1,001	1,591	591	hypothetical protein	
IYC_00442	1,762	2,730	969	plasmid segregation protein ParM	
IYC_00447	2,743	3,117	375	hypothetical protein	
IYC_00452	3,402	3,551	150	hypothetical protein	
IYC_00457	3,551	4,534	984	plasmid recombination enzyme	
IYC_00462	4,609	5,973	1,365	TraG/TraD family protein	COG3505 Type IV secretory pathway, VirD4 components
IYC_00467	6,488	7,426	939	putative phage lysozyme	COG3772 Phage-related lysozyme (muraminidase)
IYC_00472	7,461	7,877	417	toxin secretion/phage lysis holin	COG4824 Phage-related holin (Lysis protein)
IYC_00477	7,960	8,499	540	hypothetical protein	
IYC_00482	8,619	8,999	381	hypothetical protein	COG4231 Indolepyruvate ferredoxin oxidoreductase, alpha and beta subunits
IYC_00487	9,173	9,352	180	hypothetical protein	
IYC_00492	9,409	9,783	375	hypothetical protein	COG0840 Methyl-accepting chemotaxis protein
IYC_00497	9,800	11,395	1,596	hypothetical protein	
IYC_00502	11,395	12,135	741	hypothetical protein	
IYC_00507	12,122	12,847	726	hypothetical protein	
IYC_00512	12,805	13,980	1,176	putative bacteriophage protein	COG3299 Uncharacterized homolog of phage Mu proteingp47
IYC_00517	13,995	14,339	345	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_00522	14,329	14,901	573	putative bacteriophage protein	
IYC_00527	14,913	16,805	1,893	putative hydrolase	COG0791 Cell wall-associated hydrolases (invasion-associated proteins)
IYC_00532	16,883	17,206	324	hypothetical protein	
IYC_00537	17,222	17,746	525	hypothetical protein	COG1126 ABC-type polar amino acid transport system,ATPase component
IYC_00542	17,747	19,624	1,878	hypothetical protein	
IYC_00547	19,625	19,756	132	hypothetical protein	
IYC_00552	19,789	20,244	456	hypothetical protein	COG0050 GTPases - translation elongation factors
IYC_00557	20,256	20,684	429	hypothetical protein	
IYC_00562	20,699	21,802	1,104	hypothetical protein	
IYC_00567	21,820	22,374	555	hypothetical protein	
IYC_00572	22,376	22,744	369	hypothetical protein	COG1337 Uncharacterized protein predicted to be involved in DNA repair (RAMP superfamily)
IYC_00577	22,744	23,319	576	putative bacteriophage protein	
IYC_00582	23,332	23,793	462	putative bacteriophage protein	
IYC_00587	23,802	24,068	267	hypothetical protein	
IYC_00592	24,139	25,170	1,032	hypothetical protein	COG4834 Uncharacterized protein conserved in bacteria
IYC_00597	25,183	25,713	531	hypothetical protein	
IYC_00602	25,715	27,025	1,311	hypothetical protein	COG3566 Uncharacterized protein conserved in bacteria
IYC_00607	27,059	27,223	165	hypothetical protein	
IYC_00612	27,302	27,667	366	putative phage-like protein	COG2369 Uncharacterized protein, homolog of phage Mu protein gp30
IYC_00617	27,845	28,090	246	hypothetical protein	
IYC_00622	28,087	28,446	360	hypothetical protein	
IYC_00627	28,457	29,302	846	head morphogenesis protein SPP1 gp7	COG2369 Uncharacterized protein, homolog of phage Mu protein gp30
IYC_00632	29,317	29,478	162	hypothetical protein	
IYC_00637	29,471	30,949	1,479	phage-associated protein, HI1409 family	COG3567 Uncharacterized protein conserved in bacteria
IYC_00642	31,405	31,548	144	hypothetical protein	
IYC_00647	31,575	32,903	1,329	phage terminase large subunit	
IYC_00652	32,887	33,672	786	phage-like terminase small subunit yqas	
IYC_00657	33,745	33,945	201	hypothetical protein	
IYC_00662	34,057	34,518	462	hypothetical protein	
IYC_00667	34,644	35,480	837	phage integrase family protein	COG4974 Site-specific recombinase XerD
IYC_00672	35,623	36,567	945	integrase/recombinase XerD	COG0582 Integrase
IYC_00677	36,992	37,156	165	hypothetical protein	
IYC_00682	37,201	38,238	1,038	hypothetical protein	
IYC_00687	38,260	38,655	396	single-strand binding protein	COG0629 Single-stranded DNA-binding protein
IYC_00692	38,656	38,829	174	hypothetical protein	
IYC_00697	39,020	39,391	372	hypothetical protein	
IYC_00702	39,395	39,664	270	hypothetical protein	
IYC_00707	39,682	39,957	276	hypothetical protein	
IYC_00712	39,995	40,336	342	hypothetical protein	
IYC_00717	40,342	40,542	201	hypothetical protein	
IYC_00722	40,647	40,793	147	hypothetical protein	
IYC_00727	40,826	41,050	225	hypothetical protein	
IYC_00732	41,314	41,490	177	hypothetical protein	
IYC_00737	41,515	41,718	204	hypothetical protein	
IYC_00742	41,757	42,074	318	hypothetical protein	
IYC_00747	42,112	42,354	243	hypothetical protein	
IYC_00752	42,403	42,594	192	hypothetical protein	
IYC_00757	42,652	43,371	720	hypothetical protein	
IYC_00762	43,419	43,550	132	hypothetical protein	
IYC_00767	43,705	43,995	291	hypothetical protein	
IYC_00772	44,088	44,267	180	hypothetical protein	
IYC_00777	44,264	44,503	240	hypothetical protein	
IYC_00782	44,518	44,862	345	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_00787	44,853	45,077	225	hypothetical protein	
IYC_00792	45,055	>45280	>226	hypothetical protein	
Genomic Scaffold: NZ_JH470461					
IYC_00797	<1	188	>188	hypothetical protein	COG0527 Aspartokinases
IYC_00802	205	1,485	1,281	replicative DNA helicase	COG0305 Replicative DNA helicase
IYC_00807	1,501	2,247	747	hypothetical protein	
IYC_00812	2,360	2,509	150	hypothetical protein	
Genomic Scaffold: NZ_JH470462					
IYC_00817	150	263	114	hypothetical protein	
Genomic Scaffold: NZ_JH470463					
IYC_00822	291	452	162	hypothetical protein	
IYC_00827	554	1,105	552	hypothetical protein	
IYC_00832	1,253	1,498	246	hypothetical protein	
IYC_00837	1,474	1,665	192	hypothetical protein	
IYC_00842	1,843	1,983	141	hypothetical protein	
Genomic Scaffold: NZ_JH470464					
IYC_00847	<1	776	>776	N-acetylmuramoyl-L-alanine amidase	COG0860 N-acetylmuramoyl-L-alanine amidase
Genomic Scaffold: NZ_JH470465					
IYC_00852	563	1,891	1,329	AAA family ATPase	COG0714 MoxR-like ATPases
IYC_00857	1,914	3,650	1,737	von Willebrand factor type A domain-containing protein	COG2425 Uncharacterized protein containing a von Willebrand factor type A (vWA) domain
IYC_00862	3,662	4,003	342	hypothetical protein	
IYC_00867	4,040	4,768	729	BRCT domain-containing protein	COG0272 NAD-dependent DNA ligase (contains BRCT domain type II)
IYC_00872	5,365	5,934	570	DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_00877	5,984	6,886	903	glycyl-radical enzyme activating family protein	COG1180 Pyruvate-formate lyase-activating enzyme
IYC_00882	6,893	9,253	2,361	formate acetyltransferase	COG1882 Pyruvate-formate lyase
IYC_00887	9,368	10,174	807	pyrroline-5-carboxylate reductase	COG0345 Pyrroline-5-carboxylate reductase
IYC_00892	10,302	10,580	279	hypothetical protein	COG3123 Uncharacterized protein conserved in bacteria
IYC_00897	10,622	11,497	876	hypothetical protein	
IYC_00902	11,531	12,226	696	ErfK/YbiS/YcfS/YnhG family protein	COG1376 Uncharacterized protein conserved in bacteria
IYC_00907	12,285	12,872	588	putative flavin reductase	COG1853 Conserved protein/domain typically associated with flavoprotein oxygenases, DIM6/NTAB family
IYC_00912	12,910	13,413	504	hypothetical protein	COG4708 Predicted membrane protein
IYC_00917	13,634	14,863	1,230	peptidase, M16 family protein	COG0612 Predicted Zn-dependent peptidases
IYC_00922	14,885	15,820	936	recombination regulator RecX	COG2137 Uncharacterized protein conserved in bacteria
IYC_00927	15,845	17,020	1,176	transglutaminase domain-containing protein	COG1305 Transglutaminase-like enzymes, putative cysteine proteases
IYC_00932	17,151	17,498	348	TPR domain-containing protein	COG0457 FOG: TPR repeat
IYC_00937	17,642	18,445	804	ABC transporter, permease protein	COG0600 ABC-type nitrate/sulfonate/bicarbonate transport system, permease component
IYC_00942	18,442	19,203	762	ABC transporter, ATP-binding protein	COG1116 ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component
IYC_00947	19,515	23,807	4,293	putative CoA-substrate-specific enzyme activase	COG1924 Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)
IYC_00952	24,160	25,188	1,029	putative ABC transporter, substrate-binding protein	COG0715 ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components
IYC_00957	25,151	25,309	159	hypothetical protein	
IYC_00962	25,584	28,079	2,496	putative ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_00967	28,082	28,828	747	ABC transporter, ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_00972	29,266	29,772	507	GNAT family acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_00977	29,793	32,438	2,646	valyl-tRNA synthetase	COG0525 Valyl-tRNA synthetase
IYC_00982	32,692	33,981	1,290	FolC bifunctional protein	COG0285 Folylpolyglutamate synthase
IYC_00987	34,003	34,251	249	hypothetical protein	
IYC_00992	34,391	35,275	885	hypothetical protein	COG3391 Uncharacterized conserved protein

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_00997	35,278	35,805	528	hypothetical protein	
IYC_01002	35,969	36,721	753	polysaccharide deacetylase family protein	COG0726 Predicted xylanase/chitin deacetylase
IYC_01007	36,920	37,417	498	hypothetical protein	
IYC_01012	37,495	38,208	714	single-stranded DNA-binding protein	COG0629 Single-stranded DNA-binding protein
IYC_01017	38,325	39,035	711	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase	COG2171 Tetrahydrodipicolinate N-succinyltransferase
IYC_01022	39,303	40,472	1,170	aminotransferase, classes I and II	COG0436 Aspartate/tyrosine/aromatic aminotransferase
Genomic Scaffold: NZ_JH470466					
IYC_01027	<1	603	>603	hypothetical protein	
IYC_01032	665	931	267	hypothetical protein	
IYC_01037	967	>1067	>101	peptidoglycan-binding domain 1	
Genomic Scaffold: NZ_JH470467					
IYC_01042	329	925	597	uracil-DNA glycosylase family protein	COG1573 Uracil-DNA glycosylase
Genomic Scaffold: NZ_JH470468					
IYC_01049	61	282	222	DNA-binding protein	COG1396 Predicted transcriptional regulators
Genomic Scaffold: NZ_JH470469					
IYC_01054	10	1,305	1,296	hypothetical protein	COG2826 Transposase and inactivated derivatives, IS30 family
Genomic Scaffold: NZ_JH470470					
IYC_01059	<1	>368	>368	antirepressor, phage associated protein	COG3617 Prophage antirepressor
Genomic Scaffold: NZ_JH470471					
IYC_01064	105	242	138	hypothetical protein	
IYC_01069	272	364	93	hypothetical protein	
IYC_01074	390	>491	>102	antirepressor, phage associated protein	
Genomic Scaffold: NZ_JH470472					
IYC_01079	<1	>287	>287	hypothetical protein	
Genomic Scaffold: NZ_JH470473					
IYC_01084	90	1,166	1,077	hypothetical protein	COG1915 Uncharacterized conserved protein
IYC_01089	1,235	>2319	>1085	hypothetical protein	COG1915 Uncharacterized conserved protein
Genomic Scaffold: NZ_JH470474					
IYC_01094	181	855	675	hypothetical protein	COG0671 Membrane-associated phospholipid phosphatase
IYC_01099	801	1,499	699	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_01104	1,489	3,003	1,515	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_01109	3,267	4,682	1,416	sodium:solute symporter family protein	COG0591 Na+/proline symporter
IYC_01114	4,774	5,442	669	DNA-binding transcriptional activator YeiL	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_01119	5,539	6,177	639	Methyltransferase type 11	COG0500 SAM-dependent methyltransferases
IYC_01124	7,234	8,553	1,320	electron transport complex, RnfABCDGE type, C subunit	COG4656 Predicted NADH:ubiquinone oxidoreductase, subunit RnfC
IYC_01129	8,576	9,508	933	electron transport complex, RnfABCDGE type, D subunit	COG4658 Predicted NADH:ubiquinone oxidoreductase, subunit RnfD
IYC_01134	9,508	10,080	573	electron transport complex, RnfABCDGE type, G subunit	COG4659 Predicted NADH:ubiquinone oxidoreductase, subunit RnfG
IYC_01139	10,080	10,679	600	electron transport complex RsxE subunit	COG4660 Predicted NADH:ubiquinone oxidoreductase, subunit RnfE
IYC_01144	10,731	11,309	579	electron transport complex, RnfABCDGE type, A subunit	COG4657 Predicted NADH:ubiquinone oxidoreductase, subunit RnfA
IYC_01149	11,323	12,156	834	ferredoxin	COG2878 Predicted NADH:ubiquinone oxidoreductase, subunit RnfB
IYC_01154	12,570	13,628	1,059	amidase domain-containing protein	COG2247 Putative cell wall-binding domain
IYC_01159	13,768	14,304	537	hypothetical protein	COG4769 Predicted membrane protein
IYC_01164	14,322	14,714	393	hypothetical protein	COG5341 Uncharacterized protein conserved in bacteria
IYC_01169	14,831	15,865	1,035	apbE family protein	COG1477 Membrane-associated lipoprotein involved inthiamine biosynthesis

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_01174	16,512	21,101	4,590	putative surface/cell-adhesion protein/N-acetylmuramoyl-L-alanine amidase	COG2247 Putative cell wall-binding domain
IYC_01179	21,398	21,493	96	hypothetical protein	
IYC_01184	21,521	21,616	96	hypothetical protein	
IYC_01189	22,009	24,027	2,019	amidase domain-containing protein	COG2247 Putative cell wall-binding domain
IYC_01194	24,112	25,857	1,746	putative cell wall-binding protein	COG2247 Putative cell wall-binding domain
IYC_01199	26,153	32,185	6,033	putative cell surface protein	COG2247 Putative cell wall-binding domain
IYC_01204	32,689	34,203	1,515	hypothetical protein	COG5279 Uncharacterized protein involved in cytokinesis, contains TGc (transglutaminase/protease-like) domain
IYC_01209	34,626	35,537	912	ABC transporter, solute-binding protein	COG0226 ABC-type phosphate transport system, periplasmic component
IYC_01214	35,744	36,607	864	phosphate ABC transporter permease	COG0573 ABC-type phosphate transport system, permease component
IYC_01219	36,643	37,473	831	phosphate ABC transporter permease	COG0581 ABC-type phosphate transport system, permease component
IYC_01224	37,475	38,221	747	phosphate import ATP-binding protein PstB	COG1117 ABC-type phosphate transport system, ATPasecomponent
IYC_01229	38,414	38,884	471	hypothetical protein	COG4087 Soluble P-type ATPase
IYC_01234	39,444	40,895	1,452	amino acid permease family protein	COG0531 Amino acid transporters
IYC_01239	41,559	43,034	1,476	amino acid permease	COG0531 Amino acid transporters
IYC_01244	43,058	43,561	504	putative lipoprotein	
IYC_01249	43,741	44,565	825	hypothetical protein	
IYC_01254	44,644	45,339	696	hypothetical protein	COG1451 Predicted metal-dependent hydrolase
IYC_01259	45,426	45,764	339	single-strand binding domain-containing protein	COG0629 Single-stranded DNA-binding protein
IYC_01264	46,002	46,553	552	hypothetical protein	COG1971 Predicted membrane protein
IYC_01269	46,617	47,051	435	Fur family transcriptional regulator	COG0735 Fe2+/Zn2+ uptake regulation proteins
IYC_01274	47,048	47,395	348	hypothetical protein	
IYC_01279	47,667	48,719	1,053	HD domain-containing protein	COG2206 HD-GYP domain
IYC_01284	48,858	49,889	1,032	S54 family peptidase	COG0705 Uncharacterized membrane protein (homolog of Drosophila rhomboid)
IYC_01289	50,194	52,248	2,055	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_01294	52,981	54,711	1,731	ATP-dependent metalloprotease	COG0465 ATP-dependent Zn proteases
IYC_01299	54,874	55,023	150	hypothetical protein	
IYC_01304	54,932	55,990	1,059	hypothetical protein	COG3339 Uncharacterized conserved protein
IYC_01309	55,965	56,414	450	ribonuclease H	COG0328 Ribonuclease HI
IYC_01314	56,477	56,653	177	hypothetical protein	
IYC_01319	56,661	56,867	207	hypothetical protein	
IYC_01324	56,836	58,140	1,305	tetratricopeptide repeat/DNA binding domain-containing protein	COG1396 Predicted transcriptional regulators
IYC_01329	58,172	59,425	1,254	tetratricopeptide repeat protein	COG1396 Predicted transcriptional regulators
IYC_01334	59,812	59,982	171	hypothetical protein	
IYC_01339	60,270	62,870	2,601	clpB protein	COG0542 ATPases with chaperone activity, ATP-binding subunit
Genomic Scaffold: NZ_JH470475					
IYC_01344	326	1,180	855	undecaprenyl pyrophosphate phosphatase	COG1968 Uncharacterized bacitracin resistance protein
IYC_01349	1,287	2,402	1,116	putative lipoprotein	
IYC_01354	2,865	3,845	981	mannose-6-phosphate isomerase, class I	COG1482 Phosphomannose isomerase
IYC_01359	4,177	5,538	1,362	putative UDP-N-acetylmuramyl tripeptide synthetase	COG0769 UDP-N-acetylmuramyl tripeptide synthase
IYC_01364	5,598	6,326	729	CobB/CobQ family glutamine amidotransferase	COG3442 Predicted glutamine amidotransferase
Genomic Scaffold: NZ_JH470476					

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_01369	37	2,106	2,070	elongation factor G	COG0480 Translation elongation factors (GTPases)
IYC_01374	2,179	2,649	471	30S ribosomal protein S7	COG0049 Ribosomal protein S7
IYC_01379	2,810	3,187	378	30S ribosomal protein S12	COG0048 Ribosomal protein S12
IYC_01384	3,263	3,502	240	ribosomal protein L7Ae family protein	COG1358 Ribosomal protein HS6-type (S12/L30/L7a)
IYC_01389	3,627	7,163	3,537	DNA-directed RNA polymerase subunit beta'	COG0086 DNA-directed RNA polymerase, beta' subunit/160 kD subunit
IYC_01394	7,184	10,900	3,717	DNA-directed RNA polymerase subunit beta	COG0085 DNA-directed RNA polymerase, beta subunit/140 kD subunit
IYC_01399	11,165	11,536	372	50S ribosomal protein L7/L12	COG0222 Ribosomal protein L7/L12
IYC_01404	11,584	12,084	501	50S ribosomal protein L10	COG0244 Ribosomal protein L10
IYC_01409	12,287	12,976	690	50S ribosomal protein L1	COG0081 Ribosomal protein L1
IYC_01414	13,033	13,458	426	50S ribosomal protein L11	COG0080 Ribosomal protein L11
IYC_01419	13,529	14,053	525	transcription antitermination protein NusG	COG0250 Transcription antiterminator
IYC_01424	14,100	14,327	228	preprotein translocase subunit SecE	COG0690 Preprotein translocase subunit SecE
IYC_01429	14,377	14,526	150	50S ribosomal protein L33	COG0267 Ribosomal protein L33
Genomic Scaffold: NZ_JH470477					
IYC_01434	<1	146	>146	transposase	COG2801 Transposase and inactivated derivatives
IYC_01439	219	>370	>152	hypothetical protein	
Genomic Scaffold: NZ_JH470478					
IYC_01444	<1	>204	>204	transposase	COG2963 Transposase and inactivated derivatives
Genomic Scaffold: NZ_JH470479					
IYC_01449	<1	>389	>389	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter
Genomic Scaffold: NZ_JH470480					
IYC_01454	<1	127	>127	hypothetical protein	
IYC_01459	302	>1242	>941	beta-N-acetylglucosaminidase	COG4193 Beta- N-acetylglucosaminidase
Genomic Scaffold: NZ_JH470481					
IYC_01464	<1	193	>193	hypothetical protein	
IYC_01469	345	1,217	873	NAD-dependent epimerase/dehydratase	COG0451 Nucleoside-diphosphate-sugar epimerases; overlaps another CDS with the same product name
IYC_01474	1,214	2,176	963	NAD-dependent epimerase/dehydratase	COG0451 Nucleoside-diphosphate-sugar epimerases; overlaps another CDS with the same product name
IYC_01479	2,181	4,238	2,058	glycosyl transferase family protein	COG0438 Glycosyltransferase
IYC_01484	4,800	8,201	3,402	hypothetical protein	
IYC_01489	8,320	9,627	1,308	hypothetical protein	
IYC_01494	9,882	9,977	96	hypothetical protein	
IYC_01499	9,982	11,481	1,500	extracellular exochitinase	COG3469 Chitinase
IYC_01504	11,710	13,758	2,049	hypothetical protein	COG1331 Highly conserved protein containing a thioredoxin domain
IYC_01509	14,024	15,004	981	hypothetical protein	COG4129 Predicted membrane protein
IYC_01514	15,169	15,753	585	hypothetical protein	COG0515 Serine/threonine protein kinase
IYC_01519	16,065	16,766	702	CAAX amino terminal protease family protein	COG1266 Predicted metal-dependent membrane protease
IYC_01524	16,880	17,686	807	metallo-beta-lactamase family protein	COG1237 Metal-dependent hydrolases of the beta-lactamase superfamily II
IYC_01529	18,160	20,079	1,920	drug resistance ABC transporter, ATP-binding protein	COG0488 ATPase components of ABC transporters with duplicated ATPase domains
IYC_01534	20,715	21,929	1,215	hypothetical protein	COG3314 Uncharacterized protein conserved in bacteria
IYC_01539	22,298	22,636	339	ArsC family protein	COG1393 Arsenate reductase and related proteins, glutaredoxin family
IYC_01544	22,831	23,709	879	radical SAM domain-containing protein	COG2896 Molybdenum cofactor biosynthesis enzyme
IYC_01549	24,047	25,465	1,419	FAD-dependent oxidoreductase	COG0579 Predicted dehydrogenase
IYC_01554	25,521	26,777	1,257	pyridine nucleotide-disulfide oxidoreductase	COG0446 Uncharacterized NAD(FAD)-dependent dehydrogenases

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_01559	26,780	27,148	369	hypothetical protein	COG3862 Uncharacterized protein with conserved CXXCpairs
IYC_01564	27,236	28,573	1,338	glycerol-3-phosphate transporter	COG2271 Sugar phosphate permease
IYC_01569	29,170	30,861	1,692	arginyl-tRNA synthetase	COG0018 Arginyl-tRNA synthetase
IYC_01574	31,521	32,651	1,131	transaminase	COG0436 Aspartate/tyrosine/aromatic aminotransferase
IYC_01579	32,951	33,331	381	CBS domain-containing protein	COG0517 FOG: CBS domain
IYC_01584	33,509	33,673	165	hypothetical protein	
IYC_01589	33,699	35,462	1,764	ferrous iron transport protein B	COG0370 Fe2+ transport system protein B
IYC_01594	35,477	35,698	222	ferrous ion transport protein A	COG1918 Fe2+ transport system protein A
IYC_01599	36,024	36,419	396	hypothetical protein	
IYC_01604	36,435	37,160	726	haloacid dehalogenase	COG0560 Phosphoserine phosphatase
IYC_01609	37,853	39,436	1,584	putative glycosyl hydrolase	COG3979 Uncharacterized protein contain chitin-binding domain type 3
IYC_01614	39,752	40,708	957	methionine ABC transporter ATP-binding protein	COG1135 ABC-type metal ion transport system, ATPasecomponent
IYC_01619	40,710	41,351	642	methionine ABC transporter permease	COG2011 ABC-type metal ion transport system, permease component
IYC_01624	41,409	42,221	813	NLPA family lipoprotein	COG1464 ABC-type metal ion transport system, periplasmic component/surface antigen
IYC_01629	42,408	43,220	813	NLPA family lipoprotein	COG1464 ABC-type metal ion transport system, periplasmic component/surface antigen
IYC_01634	43,313	44,479	1,167	amidohydrolase family protein	COG1473 Metal-dependent amidase/aminoacylase/carboxypeptidase
IYC_01639	45,052	45,681	630	hypothetical protein	
IYC_01644	45,708	46,229	522	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_01649	46,385	47,077	693	GntR family transcriptional regulator	COG2186 Transcriptional regulators
IYC_01654	47,112	47,978	867	hypothetical protein	COG1284 Uncharacterized conserved protein
IYC_01659	48,159	48,371	213	hypothetical protein	
IYC_01664	48,586	50,043	1,458	glutamyl-tRNA synthetase	COG0008 Glutamyl- and glutaminyl-tRNA synthetases
IYC_01669	50,456	51,472	1,017	glyceraldehyde-3-phosphate dehydrogenase, type I	COG0057 Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase
IYC_01674	51,702	53,210	1,509	metallo-beta-lactamase family protein	COG1236 Predicted exonuclease of the beta-lactamasefold involved in RNA processing
IYC_01679	53,599	53,748	150	hypothetical protein	
IYC_01684	53,890	54,534	645	HD domain-containing protein	COG0617 tRNA nucleotidyltransferase/poly(A) polymerase
IYC_01689	54,590	55,960	1,371	aldehyde dehydrogenase family protein	COG1012 NAD-dependent aldehyde dehydrogenases
IYC_01694	56,041	57,438	1,398	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter
IYC_01699	58,000	59,382	1,383	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter
IYC_01704	59,816	60,256	441	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_01709	60,274	61,647	1,374	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_01714	61,878	62,420	543	TPR domain-containing protein	COG0457 FOG: TPR repeat
IYC_01719	62,523	63,500	978	hypothetical protein	
IYC_01724	64,354	65,592	1,239	hypothetical protein	COG0834 ABC-type amino acid transport/signal transduction systems, periplasmic component/domain
IYC_01729	65,594	66,466	873	ABC transporter, permease protein	COG1175 ABC-type sugar transport systems, permease components
IYC_01734	66,485	67,291	807	ABC transporter permease	COG1177 ABC-type spermidine/putrescine transport system, permease component II
IYC_01739	67,303	68,322	1,020	ABC transporter, ATP-binding protein	COG3842 ABC-type spermidine/putrescine transport systems, ATPase components
IYC_01744	68,772	69,731	960	hypothetical protein	COG4866 Uncharacterized conserved protein

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_01749	69,819	70,370	552	GNAT family acetyltransferase	COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins
IYC_01754	70,380	70,715	336	hypothetical protein	COG0526 Thiol-disulfide isomerase and thioredoxins
IYC_01759	71,287	72,270	984	birA bifunctional protein	COG1654 Biotin operon repressor
IYC_01764	72,434	72,637	204	putative heavy metal-binding protein	COG2608 Copper chaperone
IYC_01769	72,914	73,351	438	CBS domain-containing protein	COG0517 FOG: CBS domain
IYC_01784	75,051	75,170	120	hypothetical protein	
IYC_01789	75,416	75,514	99	hypothetical protein	
IYC_01794	75,790	76,569	780	hypothetical protein	
IYC_01799	77,070	78,404	1,335	magnesium transporter	COG2239 Mg/Co/Ni transporter MgtE (contains CBS domain)
IYC_01804	79,287	80,360	1,074	aminotransferase, class V	COG0075 Serine-pyruvate aminotransferase/archaeal aspartate aminotransferase
IYC_01809	80,566	81,510	945	D-3-phosphoglycerate dehydrogenase	COG0111 Phosphoglycerate dehydrogenase and related dehydrogenases
IYC_01814	81,488	82,816	1,329	hypothetical protein	COG4198 Uncharacterized conserved protein
IYC_01819	83,458	84,648	1,191	metallo-beta-lactamase/flavodoxin domain-containing protein	COG0426 Uncharacterized flavoproteins
IYC_01824	84,815	89,104	4,290	putative CoA-substrate-specific enzyme activase	COG1924 Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)
IYC_01829	89,178	90,692	1,515	choline/carnitine/betaine transporter	COG1292 Choline-glycine betaine transporter
IYC_01834	91,026	91,652	627	GntR family transcriptional regulator	COG2188 Transcriptional regulators
IYC_01839	91,945	92,295	351	hypothetical protein	COG0640 Predicted transcriptional regulators
IYC_01844	92,411	93,427	1,017	hypothetical protein	COG0701 Predicted permeases
IYC_01849	93,464	93,691	228	hypothetical protein	COG0526 Thiol-disulfide isomerase and thioredoxins
IYC_01854	93,941	94,441	501	hypothetical protein	COG1853 Conserved protein/domain typically associated with flavoprotein oxygenases, DIM6/NTAB family
IYC_01859	94,822	95,133	312	putative phage shock protein	COG0607 Rhodanese-related sulfurtransferase
IYC_01864	95,209	95,673	465	hypothetical protein	
IYC_01869	95,951	96,151	201	hypothetical protein	
IYC_01874	96,343	97,257	915	TPR repeat-containing protein	COG0457 FOG: TPR repeat
IYC_01879	97,264	97,836	573	xanthine phosphoribosyltransferase	COG0503 Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins
IYC_01884	98,137	98,718	582	hypothetical protein	
IYC_01889	99,028	100,410	1,383	putative thiosulfate sulfurtransferase	COG2897 Rhodanese-related sulfurtransferase
IYC_01894	100,609	101,031	423	hypothetical protein	
IYC_01899	101,370	102,470	1,101	putative transporter	COG0628 Predicted permease
IYC_01904	102,514	104,220	1,707	putative methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_01909	104,393	104,719	327	MazG family protein	COG1694 Predicted pyrophosphatase
IYC_01914	104,859	106,598	1,740	hypothetical protein	COG1866 Phosphoenolpyruvate carboxykinase (ATP)
IYC_01919	107,134	107,274	141	hypothetical protein	
IYC_01924	107,456	108,655	1,200	hypothetical protein	
IYC_01929	109,049	110,095	1,047	ABC transporter, substrate-binding protein	COG1840 ABC-type Fe3+ transport system, periplasmiccomponent
IYC_01934	110,097	111,785	1,689	ABC transporter, permease protein	COG1178 ABC-type Fe3+ transport system, permease component
IYC_01939	111,748	112,686	939	ABC transporter, ATP-binding protein	COG3842 ABC-type spermidine/putrescine transport systems, ATPase components
IYC_01944	112,999	114,162	1,164	30S ribosomal protein S1	COG0539 Ribosomal protein S1
IYC_01949	114,271	114,690	420	thioesterase family protein	COG0824 Predicted thioesterase
IYC_01954	114,876	115,292	417	hypothetical protein	
IYC_01959	116,183	117,190	1,008	putative ABC transporter, substrate-binding protein	COG2984 ABC-type uncharacterized transport system, periplasmic component
IYC_01964	117,190	118,095	906	ABC transporter, permease protein	COG4120 ABC-type uncharacterized transport system, permease component

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_01969	118,058	118,852	795	ABC transporter, ATP-binding protein	COG1101 ABC-type uncharacterized transport system, ATPase component
IYC_01974	119,190	119,492	303	hypothetical protein	
IYC_01979	119,924	121,567	1,644	aspartate aminotransferase	COG0436 Aspartate/tyrosine/aromatic aminotransferase
IYC_01984	121,720	122,601	882	hypothetical protein	COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily
IYC_01989	122,498	122,626	129	hypothetical protein	
IYC_01994	123,033	124,130	1,098	PfkB family kinase	COG2771 DNA-binding HTH domain-containing proteins
IYC_01999	124,124	125,047	924	indigoidine synthase A family protein	COG2313 Uncharacterized enzyme involved in pigment biosynthesis
IYC_02004	125,424	126,335	912	2-dehydropantoate 2-reductase	COG1893 Ketopantoate reductase
IYC_02009	126,520	127,152	633	methenyltetrahydrofolate cyclohydrolase	COG3404 Methenyl tetrahydrofolate cyclohydrolase
IYC_02014	127,152	128,000	849	bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase	COG0190 5,10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate cyclohydrolase
IYC_02019	128,555	129,349	795	HAD hydrolase	COG0561 Predicted hydrolases of the HAD superfamily
IYC_02024	129,681	130,187	507	putative diamine N-acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_02029	130,380	130,523	144	hypothetical protein	
IYC_02034	130,643	131,380	738	hypothetical protein	
IYC_02039	131,494	132,303	810	ZIP family zinc transporter	COG0428 Predicted divalent heavy-metal cations transporter
IYC_02044	132,974	133,411	438	rrf2 family protein	COG1959 Predicted transcriptional regulator
IYC_02049	133,745	134,674	930	cysteine synthase A	COG0031 Cysteine synthase
IYC_02054	134,718	135,386	669	His/Glu/Gln/Arg/opine ABC transporter permease	COG0765 ABC-type amino acid transport system, permease component
IYC_02059	135,421	136,242	822	polar amino acid ABC transporter amino acid-binding protein	COG0834 ABC-type amino acid transport/signal transduction systems, periplasmic component/domain
IYC_02064	136,347	137,477	1,131	cysteine desulfhydrase	COG1104 Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes
IYC_02069	137,598	138,398	801	hypothetical protein	COG1606 ATP-utilizing enzymes of the PP-loop superfamily
IYC_02074	138,547	138,675	129	hypothetical protein	
IYC_02079	138,792	139,862	1,071	tRNA-specific 2-thiouridylase MnmA	COG0482 Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase, contains the PP-loop ATPase domain
IYC_02084	140,042	140,797	756	hypothetical protein	
IYC_02089	140,862	141,458	597	hypothetical protein	
IYC_02094	141,746	142,051	306	hypothetical protein	
IYC_02099	142,299	142,868	570	hypothetical protein	
IYC_02104	143,208	143,690	483	flavodoxin	COG0716 Flavodoxins
IYC_02109	143,843	144,010	168	hypothetical protein	
IYC_02114	144,050	147,568	3,519	pyruvate ferredoxin oxidoreductase	COG0674 Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, alpha subunit
IYC_02119	147,949	148,134	186	hypothetical protein	
IYC_02124	148,727	150,202	1,476	hypothetical protein	COG0397 Uncharacterized conserved protein
IYC_02129	150,720	152,459	1,740	amine oxidase	COG1231 Monoamine oxidase
IYC_02134	152,503	153,393	891	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_02139	153,593	154,969	1,377	sugar transporter	COG2211 Na+/melibiose symporter and related transporters
IYC_02144	154,983	156,311	1,329	O-glycosyl hydrolase, family protein	COG5520 O-Glycosyl hydrolase
IYC_02149	156,375	157,316	942	ROK family protein	COG1940 Transcriptional regulator/sugar kinase
IYC_02154	157,319	158,764	1,446	glucosyl hydrolase family protein	COG2723 Beta-glucosidase/6-phospho-beta-glucosidase/beta- galactosidase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_02159	158,856	159,407	552	putative transcriptional regulatory protein	COG1396 Predicted transcriptional regulators
IYC_02164	159,645	160,232	588	hypothetical protein	COG2316 Predicted hydrolase (HD superfamily)
IYC_02169	160,363	160,473	111	hypothetical protein	
IYC_02174	160,463	161,008	546	RNA polymerase factor sigma-70	COG1595 DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
IYC_02179	161,008	162,339	1,332	hypothetical protein	
IYC_02184	162,623	163,213	591	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_02189	163,479	164,177	699	phosphonate-transporting ATPase	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_02194	164,191	167,295	3,105	hypothetical protein	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_02199	167,625	167,783	159	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_02204	167,765	167,944	180	hypothetical protein	
IYC_02209	168,134	168,871	738	putative ABC transporter permease	COG4200 Uncharacterized protein conserved in bacteria
IYC_02214	168,885	169,652	768	hypothetical protein	COG4200 Uncharacterized protein conserved in bacteria
IYC_02219	169,655	170,383	729	hypothetical protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_02224	170,603	171,322	720	putative transcriptional regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_02229	171,335	172,747	1,413	putative sensory transduction histidine kinase	COG0642 Signal transduction histidine kinase
IYC_02234	173,297	173,503	207	beta-amylase	
IYC_02239	174,156	175,358	1,203	tyrosyl-tRNA synthetase	COG0162 Tyrosyl-tRNA synthetase
IYC_02244	175,648	176,331	684	ABC transporter-like protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_02249	176,321	178,792	2,472	efflux ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_02254	178,960	179,637	678	regulatory protein VanR	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_02259	179,637	180,650	1,014	two-component system histidine kinase	COG0642 Signal transduction histidine kinase
IYC_02264	180,797	180,955	159	hypothetical protein	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_02269	181,054	181,419	366	GntR family transcriptional regulator	COG1725 Predicted transcriptional regulators
IYC_02274	181,412	182,281	870	ABC transporter, ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_02279	182,285	183,034	750	hypothetical protein	
IYC_02284	183,203	183,508	306	CAAX amino terminal protease family protein	
IYC_02289	183,556	184,011	456	acetyltransferase	COG2153 Predicted acyltransferase
IYC_02294	184,257	184,625	369	hypothetical protein	
IYC_02299	184,630	184,755	126	hypothetical protein	
IYC_02304	184,969	185,790	822	hypothetical protein	
IYC_02309	185,991	187,772	1,782	hypothetical protein	
IYC_02324	189,535	190,794	1,260	putative sensory box-containing diguanylate cyclase	COG2199 FOG: GGDEF domain
IYC_02329	191,183	191,857	675	L-serine dehydratase, iron-sulfur-dependent, beta subunit	COG1760 L-serine deaminase
IYC_02334	191,870	192,760	891	L-serine dehydratase, iron-sulfur-dependent, alpha subunit	COG1760 L-serine deaminase
IYC_02339	193,000	193,836	837	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_02344	194,065	194,274	210	hypothetical protein	
IYC_02349	194,428	194,571	144	hypothetical protein	
IYC_02354	194,841	195,425	585	hypothetical protein	COG5663 Uncharacterized conserved protein
IYC_02359	195,447	196,415	969	dipeptidase family protein	COG2355 Zn-dependent dipeptidase, microsomal dipeptidase homolog

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_02364	196,812	197,501	690	hypothetical protein	
IYC_02369	197,608	198,891	1,284	putative ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_02374	198,893	199,708	816	hypothetical protein	COG1361 S-layer domain
IYC_02379	199,745	200,419	675	putative oligopeptide ABC transporter ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_02384	200,592	201,440	849	putative oligopeptide ABC transporter, permease protein	COG1173 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components
IYC_02389	201,446	202,399	954	putative oligopeptide ABC transporter, permease protein	COG0601 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components
IYC_02394	202,598	204,181	1,584	oligopeptide ABC transporter, oligopeptide-binding protein	COG0747 ABC-type dipeptide transport system, periplasmic component
IYC_02399	204,474	204,635	162	hypothetical protein	
IYC_02404	204,669	204,941	273	spore coat peptide assembly protein CotJB	
IYC_02409	204,972	205,544	573	CotJC protein	COG3546 Mn-containing catalase
IYC_02414	205,774	206,793	1,020	hypothetical protein	COG4632 Exopolysaccharide biosynthesis protein related to N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase
IYC_02419	206,810	207,874	1,065	hypothetical protein	
IYC_02424	208,299	208,790	492	hypothetical protein	COG4894 Uncharacterized conserved protein
IYC_02429	208,964	209,356	393	hypothetical protein	
IYC_02434	209,709	210,335	627	deoxyuridine 5'-triphosphate nucleotidohydrolase	COG0756 dUTPase
IYC_02439	210,393	210,584	192	small acid-soluble spore protein, H-type	
IYC_02444	210,786	210,893	108	hypothetical protein	
IYC_02449	211,105	211,860	756	hypothetical protein	
IYC_02454	212,256	212,402	147	hypothetical protein	
IYC_02459	212,521	213,051	531	putative transcriptional regulator	COG1396 Predicted transcriptional regulators
IYC_02464	213,475	214,524	1,050	putative transcriptional regulator	COG1299 Phosphotransferase system, fructose-specific IIC component
IYC_02469	214,874	216,310	1,437	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter
IYC_02474	217,466	217,774	309	hypothetical protein	
IYC_02479	218,127	218,873	747	putative lipoprotein	COG4758 Predicted membrane protein
IYC_02484	219,192	219,419	228	rubredoxin	COG0695 Glutaredoxin and related proteins
IYC_02489	219,479	220,939	1,461	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	COG1012 NAD-dependent aldehyde dehydrogenases
IYC_02494	220,979	221,920	942	putative thioredoxin reductase	COG0492 Thioredoxin reductase
IYC_02499	222,681	223,058	378	GrdX protein	
IYC_02504	223,149	224,435	1,287	glycine reductase subunits ABC	COG1331 Highly conserved protein containing a thioredoxin domain
IYC_02519	225,947	226,888	942	thioredoxin-disulfide reductase	COG0492 Thioredoxin reductase
IYC_02524	226,923	227,243	321	thioredoxin family protein	COG0526 Thiol-disulfide isomerase and thioredoxins
IYC_02529	227,328	227,459	132	hypothetical protein	
IYC_02534	227,481	227,804	324	glycine/sarcosine/betaine reductase complex protein A	
IYC_02539	227,896	229,434	1,539	glycine reductase complex component C subunit beta	COG0150 Phosphoribosylaminoimidazole (AIR) synthetase
IYC_02544	229,450	230,616	1,167	glycine reductase complex component C subunit alpha	COG0416 Fatty acid/phospholipid biosynthesis enzyme
IYC_02549	231,027	231,638	612	3-methyladenine DNA glycosylase	COG2094 3-methyladenine DNA glycosylase
IYC_02554	231,705	233,936	2,232	restriction endonuclease family protein	COG1002 Type II restriction enzyme, methylase subunits
IYC_02559	234,221	234,460	240	N-acetylmuramoyl-L-alanine amidase	
IYC_02564	234,501	234,674	174	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_02569	235,164	236,129	966	NlpC/P60 family protein	COG0791 Cell wall-associated hydrolases (invasion-associated proteins)
IYC_02594	237,982	238,341	360	hypothetical protein	
IYC_02599	238,694	239,569	876	cation efflux family protein	COG0053 Predicted Co/Zn/Cd cation transporters
IYC_02604	239,817	240,545	729	hypothetical protein	COG2323 Predicted membrane protein
IYC_02609	240,707	241,072	366	hypothetical protein	
IYC_02614	241,204	241,542	339	hypothetical protein	
IYC_02619	241,539	242,012	474	hypothetical protein	
IYC_02624	242,265	242,837	573	transcriptional regulator, TetR family protein	COG1309 Transcriptional regulator
IYC_02629	243,429	244,007	579	hypothetical protein	
IYC_02634	244,043	244,738	696	hypothetical protein	
IYC_02639	244,803	244,976	174	hypothetical protein	
IYC_02644	245,334	245,777	444	sporulation protein YtfJ	COG3874 Uncharacterized conserved protein
IYC_02649	245,799	246,344	546	membrane associated protein	
IYC_02654	246,523	246,723	201	hypothetical protein	
IYC_02659	247,537	247,821	285	hypothetical protein	COG4095 Uncharacterized conserved protein
IYC_02664	247,859	248,155	297	hypothetical protein	COG4095 Uncharacterized conserved protein
IYC_02669	248,213	250,510	2,298	xanthine dehydrogenase subunit XdhA	COG1529 Aerobic-type carbon monoxide dehydrogenase,large subunit CoxL/CutL homologs
IYC_02674	250,524	251,414	891	xanthine dehydrogenase subunit XdhB	COG1319 Aerobic-type carbon monoxide dehydrogenase,middle subunit CoxM/CutM homologs
IYC_02679	251,451	251,924	474	xanthine dehydrogenase family protein, iron-sulfur binding subunit	COG2080 Aerobic-type carbon monoxide dehydrogenase,small subunit CoxS/CutS homologs
IYC_02684	252,040	252,513	474	guanine deaminase	COG0590 Cytosine/adenosine deaminases
IYC_02689	253,102	255,297	2,196	DNA topoisomerase III	COG0550 Topoisomerase IA
IYC_02694	255,343	256,404	1,062	permease	COG4767 Glycopeptide antibiotics resistance protein
IYC_02699	256,673	257,278	606	hypothetical protein	COG1971 Predicted membrane protein
IYC_02704	257,336	257,572	237	hypothetical protein	
IYC_02709	258,122	258,808	687	NPQTN cell wall surface anchor protein	COG5386 Cell surface protein
IYC_02714	258,886	259,779	894	iron (III) dicitrate-binding periplasmic protein	COG0614 ABC-type Fe3+-hydroxamate transport system,periplasmic component
IYC_02719	259,800	260,795	996	iron compound ABC transporter permease	COG0609 ABC-type Fe3+-siderophore transport system,permease component
IYC_02724	260,788	261,564	777	ferric enterobactin transport ATP-binding protein fepC	COG1120 ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components
IYC_02729	261,596	262,204	609	sortase	COG3764 Sortase (surface protein transpeptidase)
IYC_02734	262,432	265,578	3,147	hypothetical protein	COG0206 Cell division GTPase
IYC_02739	265,995	266,648	654	putative translaldolase	COG0176 Transaldolase
IYC_02744	267,119	268,783	1,665	RNA-metabolising metallo-beta-lactamase	COG0595 Predicted hydrolase of the metallo-beta-lactamase superfamily
IYC_02749	269,033	>269825	>793	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter
Genomic Scaffold: NZ_JH470482					
IYC_02754	<1	>296	>296	putative electron transport protein	COG4656 Predicted NADH:ubiquinone oxidoreductase, subunit RnfC
Genomic Scaffold: NZ_JH470483					
IYC_02759	249	1,511	1,263	sodium:dicarboxylate symporter family protein	COG1301 Na+/H+-dicarboxylate symporters
IYC_02764	1,671	1,931	261	hypothetical protein	COG0526 Thiol-disulfide isomerase and thioredoxins
IYC_02769	1,931	3,022	1,092	putative permease	COG0701 Predicted permeases
IYC_02774	3,038	3,502	465	putative lipoprotein	
IYC_02779	3,520	4,620	1,101	putative lipoprotein	COG0715 ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components
IYC_02784	4,648	5,214	567	hypothetical protein	
IYC_02789	5,221	5,361	141	hypothetical protein	
IYC_02794	5,370	5,588	219	hypothetical protein	
IYC_02799	5,640	6,176	537	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_02804	6,125	6,766	642	ABC transporter permease	COG0600 ABC-type nitrate/sulfonate/bicarbonate transport system, permease component
IYC_02809	6,806	6,943	138	hypothetical protein	
IYC_02814	6,943	7,701	759	nitrate transport ATP-binding protein NrtD	COG1116 ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component
IYC_02819	7,868	8,200	333	hypothetical protein	
IYC_02834	9,584	10,843	1,260	putative R-2-hydroxyglutaryl-CoA dehydratase subunit	COG1024 Enoyl-CoA hydratase/carnithine racemase
IYC_02839	11,136	11,333	198	hypothetical protein	
IYC_02844	11,356	12,114	759	(R)-hydroxyglutaryl-CoA dehydratase activator	COG1924 Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)
IYC_02849	12,370	13,293	924	aspartate carbamoyltransferase catalytic subunit	COG0540 Aspartate carbamoyltransferase, catalytic chain
IYC_02854	13,295	13,735	441	aspartate carbamoyltransferase regulatory subunit	COG1781 Aspartate carbamoyltransferase, regulatory subunit
IYC_02859	13,757	14,947	1,191	dihydroorotase, multifunctional complex type	COG0044 Dihydroorotase and related cyclic amidohydrolases
IYC_02864	14,971	15,822	852	orotidine 5'-phosphate decarboxylase	COG0284 Orotidine-5'-phosphate decarboxylase
IYC_02869	15,965	16,720	756	dihydroorotate dehydrogenase electron transfer subunit	COG0543 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases
IYC_02874	16,720	17,616	897	dihydroorotate dehydrogenase 1B	COG0167 Dihydroorotate dehydrogenase
IYC_02879	17,650	18,225	576	orotate phosphoribosyltransferase	COG0461 Orotate phosphoribosyltransferase
IYC_02884	18,444	19,256	813	hypothetical protein	
IYC_02889	19,361	20,146	786	AP endonuclease	COG1082 Sugar phosphate isomerases/epimerases
IYC_02894	20,324	21,616	1,293	trigger factor	COG0544 FKBP-type peptidyl-prolyl cis-trans isomerase (trigger factor)
IYC_02899	21,741	22,325	585	ATP-dependent Clp protease proteolytic subunit	COG0740 Protease subunit of ATP-dependent Clp proteases
IYC_02904	22,345	23,634	1,290	ATP-dependent protease ATP-binding subunit ClpX	COG1219 ATP-dependent protease Clp, ATPase subunit
IYC_02909	23,764	25,443	1,680	ATP-dependent protease LonB	COG1067 Predicted ATP-dependent protease
IYC_02914	25,581	27,899	2,319	ATP-dependent protease La	COG0466 ATP-dependent Lon protease, bacterial type
IYC_02919	27,917	28,501	585	GTP-binding protein YsxC	COG0218 Predicted GTPase
IYC_02924	28,525	29,250	726	CAAX amino terminal protease family protein	COG1266 Predicted metal-dependent membrane protease
IYC_02929	29,642	29,749	108	hypothetical protein	
IYC_02934	29,919	30,149	231	hypothetical protein	
IYC_02939	30,211	30,621	411	hypothetical protein	
IYC_02944	30,836	31,792	957	hypothetical protein	COG2378 Predicted transcriptional regulator
IYC_02949	31,952	32,887	936	hypothetical protein	
IYC_02954	32,988	34,223	1,236	extracellular solute-binding protein	COG1653 ABC-type sugar transport system, periplasmic component
IYC_02959	34,292	34,720	429	FUR family transcriptional regulator	COG0735 Fe2+/Zn2+ uptake regulation proteins
IYC_02964	34,946	35,560	615	putative lipoprotein	
IYC_02969	35,723	36,292	570	ribosomal 5S rRNA E-loop binding protein Ctc/L25/TL5	COG1825 Ribosomal protein L25 (general stress protein Ctc)
IYC_02974	36,510	36,761	252	hypothetical protein	
IYC_02979	36,932	37,093	162	hypothetical protein	
IYC_02984	37,270	39,498	2,229	formate acetyltransferase 1	COG1882 Pyruvate-formate lyase
IYC_02989	39,675	40,373	699	pyruvate formate-lyase activating enzyme	COG1180 Pyruvate-formate lyase-activating enzyme
IYC_02994	40,965	41,510	546	rubredoxin/rubrerythrin	COG1592 Rubrerythrin
IYC_03009	43,815	44,570	756	glycerophosphoryl diester phosphodiesterase family protein	COG0584 Glycerophosphoryl diester phosphodiesterase
IYC_03014	44,700	45,065	366	hypothetical protein	
IYC_03019	45,108	45,572	465	phosphodiesterase	COG0622 Predicted phosphoesterase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_03024	45,780	46,541	762	prolipoprotein diacylglyceryl transferase	COG0682 Prolipoprotein diacylglyceryltransferase
IYC_03029	46,680	46,832	153	hypothetical protein	
IYC_03034	46,901	47,797	897	HDIG domain-containing protein	COG3481 Predicted HD-superfamily hydrolase
IYC_03039	48,005	49,297	1,293	xanthine/uracil permease family protein	COG2252 Permeases
IYC_03044	49,578	50,675	1,098	GIY-YIG domain-containing protein	COG0322 Nuclease subunit of the excinuclease complex
IYC_03049	50,719	50,937	219	hypothetical protein	
IYC_03054	51,352	52,134	783	3-hydroxybutyryl-CoA dehydratase	COG1024 Enoyl-CoA hydratase/carnithine racemase
IYC_03059	52,154	52,984	831	3-hydroxybutyryl-CoA dehydrogenase	COG1250 3-hydroxyacyl-CoA dehydrogenase
IYC_03064	53,009	54,187	1,179	putative acyltransferase	COG0183 Acetyl-CoA acetyltransferase
IYC_03069	54,389	55,528	1,140	butyryl-CoA dehydrogenase	COG1960 Acyl-CoA dehydrogenases
IYC_03074	55,541	56,320	780	electron transfer flavoprotein subunit beta/FixAfamily protein	COG2086 Electron transfer flavoprotein, beta subunit
IYC_03079	56,336	57,340	1,005	electron transfer flavoprotein, alpha subunit/FixB family protein	COG2025 Electron transfer flavoprotein, alpha subunit
IYC_03084	57,700	59,325	1,626	sigma-54 dependent transcriptional regulator	COG3829 Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding domains
IYC_03089	59,410	60,384	975	putative CoA-substrate-specific enzyme activase	COG1924 Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)
IYC_03094	60,365	61,462	1,098	putative R-2-hydroxyglutaryl-CoA dehydratase subunit	COG3581 Uncharacterized protein conserved in bacteria
IYC_03099	61,465	62,358	894	hypothetical protein	COG3580 Uncharacterized protein conserved in bacteria
IYC_03104	62,610	62,804	195	iron-sulfur cluster-binding protein	COG1141 Ferredoxin
Genomic Scaffold: NZ_JH470485					
IYC_03109	32	1,225	1,194	elongation factor Tu	COG0050 GTPases - translation elongation factors
Genomic Scaffold: NZ_JH470486					
IYC_03114	<1	149	>149	hypothetical protein	
IYC_03119	151	318	168	hypothetical protein	
Genomic Scaffold: NZ_JH470487					
IYC_03124	81	692	612	hypothetical protein	
IYC_03129	872	1,465	594	Holliday junction DNA helicase RuvA	COG0632 Holliday junction resolvasome, DNA-binding subunit
IYC_03134	1,485	2,513	1,029	Holliday junction DNA helicase RuvB	COG2255 Holliday junction resolvasome, helicase subunit
IYC_03139	2,547	3,572	1,026	S-adenosylmethionine:tRNA ribosyltransferase-isomerase	COG0809 S-adenosylmethionine:tRNA-ribosyltransferase-isomerase (queuine synthetase)
IYC_03144	3,615	4,745	1,131	queuine tRNA-ribosyltransferase	COG0343 Queuine/archaeosine tRNA-ribosyltransferase
IYC_03149	4,924	5,259	336	preprotein translocase subunit YajC	COG1862 Preprotein translocase subunit YajC
IYC_03154	5,314	5,682	369	hypothetical protein	
IYC_03159	5,773	5,910	138	hypothetical protein	
IYC_03164	6,067	7,434	1,368	radical SAM domain-containing protein	COG0641 Arylsulfatase regulator (Fe-S oxidoreductase)
IYC_03169	7,526	8,794	1,269	preprotein translocase subunit SecD	COG0342 Preprotein translocase subunit SecD
IYC_03174	8,796	9,659	864	preprotein translocase subunit SecF	COG0341 Preprotein translocase subunit SecF
IYC_03179	9,889	10,767	879	DHH domain-containing protein	COG0608 Single-stranded DNA-specific exonuclease
IYC_03184	10,876	11,394	519	adenine phosphoribosyltransferase	COG0503 Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins
IYC_03189	11,519	13,690	2,172	GTP pyrophosphokinase	COG0317 Guanosine polyphosphate pyrophosphohydrolases/synthetases
IYC_03194	13,723	14,172	450	D-tyrosyl-tRNA(Tyr) deacylase	COG1490 D-Tyr-tRNAtyr deacylase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_03199	14,225	14,824	600	metallo-beta-lactamase family protein	COG0491 Zn-dependent hydrolases, including glyoxylases
IYC_03204	14,863	16,293	1,431	coproporphyrinogen III oxidase	COG0635 Coproporphyrinogen III oxidase and related Fe-S oxidoreductases
IYC_03209	16,312	17,559	1,248	histidyl-tRNA synthetase	COG0124 Histidyl-tRNA synthetase
IYC_03214	17,600	19,381	1,782	aspartyl-tRNA synthetase	COG0173 Aspartyl-tRNA synthetase
IYC_03219	19,416	19,844	429	sigma-E factor regulator	COG3086 Positive regulator of sigma E activity
IYC_03224	20,016	20,144	129	hypothetical protein	
IYC_03229	20,288	20,578	291	hypothetical protein	
IYC_03234	20,635	20,754	120	hypothetical protein	
IYC_03239	20,998	21,138	141	hypothetical protein	
IYC_03244	21,322	21,525	204	small, acid-soluble spore protein	
IYC_03249	21,964	23,082	1,119	putative phage-like integrase	COG0582 Integrase
IYC_03254	23,082	23,231	150	hypothetical protein	
IYC_03259	23,549	23,959	411	DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_03264	24,134	24,358	225	hypothetical protein	COG1476 Predicted transcriptional regulators
IYC_03269	24,496	25,299	804	BRO family protein	COG3617 Prophage antirepressor
IYC_03274	25,300	25,821	522	hypothetical protein	
IYC_03279	25,833	26,009	177	hypothetical protein	
IYC_03284	25,987	26,232	246	hypothetical protein	
IYC_03289	26,301	26,495	195	hypothetical protein	
IYC_03294	26,532	26,744	213	putative DNA-binding protein	COG1396 Predicted transcriptional regulators; overlaps another CDS with the same product name
IYC_03299	26,744	26,953	210	putative DNA-binding protein	overlaps another CDS with the same product name
IYC_03304	27,198	27,386	189	hypothetical protein	
IYC_03309	27,434	29,164	1,731	hypothetical protein	COG0419 ATPase involved in DNA repair
IYC_03314	29,176	30,225	1,050	putative YqaJ	COG5377 Phage-related protein, predicted endonuclease
IYC_03319	30,238	31,119	882	RecT	COG3723 Recombinational DNA repair protein (RecE pathway)
IYC_03324	31,122	31,268	147	hypothetical protein	
IYC_03329	31,284	31,460	177	hypothetical protein	
IYC_03334	31,613	31,846	234	hypothetical protein	
IYC_03339	31,873	32,619	747	putative replication initiation and membrane attachment	COG3935 Putative primosome component and related proteins
IYC_03344	32,616	33,929	1,314	replicative DNA helicase	COG0305 Replicative DNA helicase
IYC_03349	33,935	34,075	141	hypothetical protein	
IYC_03354	34,092	34,325	234	hypothetical protein	
IYC_03359	34,337	35,215	879	phage protein	
IYC_03364	35,217	35,543	327	hypothetical protein	
IYC_03369	35,550	35,669	120	hypothetical protein	
IYC_03374	35,696	35,968	273	hypothetical protein	COG3956 Protein containing tetrapyrrole methyltransferase domain and MazG-like (predicted pyrophosphatase) domain
IYC_03379	35,988	36,677	690	hypothetical protein	
IYC_03384	36,780	37,172	393	replication terminator protein, phage associated	
IYC_03389	37,199	37,912	714	hypothetical protein	
IYC_03394	37,989	38,354	366	phage protein	
IYC_03399	38,344	38,502	159	hypothetical protein	
IYC_03404	38,499	38,900	402	hypothetical protein	
IYC_03409	38,916	39,263	348	hypothetical protein	
IYC_03414	39,296	39,490	195	hypothetical protein	
IYC_03419	39,524	39,664	141	hypothetical protein	
IYC_03424	39,665	39,778	114	hypothetical protein	
IYC_03429	39,792	40,112	321	hypothetical protein	
Genomic Scaffold: NZ_JH470488					
IYC_03436	442	1,305	864	metallo-beta-lactamase family protein	COG2333 Predicted hydrolase (metallo-beta-lactamasesuperfamily)
IYC_03441	1,443	2,165	723	putative 1-acyl-sn-glycerol-3-phosphate acyltransferase	COG0204 1-acyl-sn-glycerol-3-phosphate acyltransferase
IYC_03446	2,701	4,062	1,362	HMGL-related enzyme	COG0119 Isopropylmalate/homocitrate/citramalate synthases
IYC_03451	4,206	6,128	1,923	aconitate hydratase	COG1048 Aconitase A

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_03456	6,165	7,163	999	isocitrate/isopropylmalate family dehydrogenase	COG0473 Isocitrate/isopropylmalate dehydrogenase
IYC_03461	7,416	8,198	783	polar amino acid ABC transporter ATP-binding protein	COG1126 ABC-type polar amino acid transport system,ATPase component
IYC_03466	8,702	10,471	1,770	Na+/H+ antiporter family protein	COG1757 Na+/H+ antiporter
IYC_03471	10,498	10,713	216	hypothetical protein	COG2155 Uncharacterized conserved protein
IYC_03476	10,802	11,185	384	aspartate alpha-decarboxylase	COG0853 Aspartate 1-decarboxylase
IYC_03481	11,300	12,145	846	pantoate--beta-alanine ligase	COG0414 Panthothenate synthetase
IYC_03486	12,380	13,207	828	3-methyl-2-oxobutanoate hydroxymethyltransferase	COG0413 Ketopantoate hydroxymethyltransferase
IYC_03491	13,912	17,355	3,444	ATP-dependent nuclease subunit B	COG3857 ATP-dependent nuclease, subunit B
IYC_03496	17,479	17,994	516	hypothetical protein	
IYC_03501	18,305	18,955	651	TPR repeat-containing protein	
IYC_03506	18,958	19,644	687	hypothetical protein	
IYC_03511	19,844	20,119	276	hypothetical protein	
IYC_03516	20,749	21,555	807	hypothetical protein	
IYC_03521	21,848	25,687	3,840	recombination helicase AddA	COG1074 ATP-dependent exoDNAse (exonuclease V) betasubunit (contains helicase and exonuclease domains)
IYC_03526	25,827	27,518	1,692	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_03531	27,594	28,820	1,227	peptidase T	COG2195 Di- and tripeptidases
IYC_03536	28,960	29,892	933	phospholipase family protein	COG1752 Predicted esterase of the alpha-beta hydrolase superfamily
IYC_03541	30,181	30,936	756	alpha/beta fold family hydrolase	COG0596 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
IYC_03546	31,149	31,373	225	hypothetical protein	
IYC_03551	31,440	32,156	717	metallo-beta-lactamase family protein	COG0491 Zn-dependent hydrolases, including glyoxylases
IYC_03556	32,395	32,853	459	glyoxalase	COG0346 Lactoylglutathione lyase and related lyases
IYC_03561	32,989	33,093	105	hypothetical protein	
IYC_03566	33,538	34,173	636	His/Glu/Gln/Arg/opine ABC transporter permease	COG0765 ABC-type amino acid transport system, permease component
IYC_03571	34,186	34,989	804	polar amino acid ABC transporter amino acid-binding protein	COG0834 ABC-type amino acid transport/signal transduction systems, periplasmic component/domain
IYC_03576	35,209	35,937	729	hypothetical protein	COG0500 SAM-dependent methyltransferases
IYC_03581	36,091	37,302	1,212	EAL domain-containing protein	COG3434 Predicted signal transduction protein containing EAL and modified HD-GYP domains
IYC_03586	37,573	38,190	618	thiamine-phosphate pyrophosphorylase	COG0352 Thiamine monophosphate synthase
IYC_03591	38,181	38,972	792	hydroxyethylthiazole kinase	COG2145 Hydroxyethylthiazole kinase, sugar kinase family
IYC_03596	39,043	39,843	801	phosphomethylpyrimidine kinase	COG0351 Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase
IYC_03601	40,360	41,262	903	D-alanine--D-alanine ligase	COG1181 D-alanine-D-alanine ligase and related ATP-grasp enzymes
IYC_03606	41,418	42,560	1,143	putative high-affinity branched-chain amino acidABC transporter, amino acid-binding protein	COG0683 ABC-type branched-chain amino acid transport systems, periplasmic component
IYC_03611	42,564	43,436	873	high-affinity branched-chain amino acid ABC transporter, permease protein	COG0559 Branched-chain amino acid ABC-type transport system, permease components
IYC_03616	43,437	44,381	945	putative high-affinity branched-chain amino acidABC transporter permease	COG4177 ABC-type branched-chain amino acid transport system, permease component
IYC_03621	44,393	45,130	738	putative high-affinity branched chain amino acidABC transporter, ATP-binding protein	COG0411 ABC-type branched-chain amino acid transport systems, ATPase component
IYC_03626	45,132	45,833	702	putative high-affinity branched chain amino acidABC transporter, ATP-binding protein	COG0410 ABC-type branched-chain amino acid transport systems, ATPase component
IYC_03631	45,862	46,638	777	hypothetical protein	COG2043 Uncharacterized protein conserved in archaea
IYC_03636	47,061	47,819	759	hypothetical protein	COG2966 Uncharacterized conserved protein
IYC_03641	47,813	48,256	444	hypothetical protein	COG3610 Uncharacterized conserved protein
IYC_03646	48,298	49,749	1,452	GntR family transcriptional regulator	COG1167 Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs
IYC_03651	49,811	50,380	570	hypothetical protein	COG1704 Uncharacterized conserved protein
IYC_03656	50,389	51,240	852	hypothetical protein	COG1512 Beta-propeller domains of methanol dehydrogenase type
IYC_03661	51,404	52,291	888	NUDIX family hydrolase	COG1051 ADP-ribose pyrophosphatase
IYC_03666	52,424	52,906	483	hypothetical protein	
IYC_03671	53,083	53,607	525	hypothetical protein	
IYC_03676	53,609	54,199	591	methyltransferase-like protein	COG2226 Methylase involved in ubiquinone/menaquinone biosynthesis
IYC_03681	54,301	54,531	231	hypothetical protein	COG3655 Predicted transcriptional regulator
IYC_03686	54,561	55,043	483	hypothetical protein	
IYC_03691	55,075	55,554	480	hypothetical protein	
Genomic Scaffold: NZ_JH470489					
IYC_03696	137	>243	>107	hypothetical protein	
Genomic Scaffold: NZ_JH470490					
IYC_03701	<1	617	>617	hypothetical protein	
IYC_03706	787	954	168	hypothetical protein	
IYC_03711	1,047	1,670	624	putative sporulation protein YtaF	
IYC_03716	2,019	2,810	792	endonuclease/exonuclease/phosphatase family protein	
IYC_03721	2,814	4,472	1,659	PTS system, glucose/glucoside family, IIBC component	COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific
IYC_03726	4,515	6,179	1,665	glucosyl hydrolase family protein	COG0366 Glycosidases
IYC_03731	6,542	7,489	948	putative glucokinase	COG1940 Transcriptional regulator/sugar kinase
IYC_03736	7,786	8,787	1,002	LacI family transcription regulator	COG1609 Transcriptional regulators
IYC_03741	9,336	11,060	1,725	ABC transporter, ATP-binding/permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_03746	11,062	12,816	1,755	ABC transporter ATP-binding/permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_03751	13,135	14,124	990	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_03756	14,396	14,851	456	hypothetical protein	
IYC_03761	15,311	16,192	882	peptidase, M23/M37 family protein	COG0739 Membrane proteins related to metalloendopeptidases
IYC_03766	16,356	17,171	816	RelA/SpoT domain-containing protein	COG2357 Uncharacterized protein conserved in bacteria
IYC_03771	17,428	20,724	3,297	helicase, Snf2 family protein	COG0553 Superfamily II DNA/RNA helicases, SNF2 family
IYC_03776	21,308	22,669	1,362	branched-chain amino acid transport system II carrier protein	COG1114 Branched-chain amino acid permeases
IYC_03781	23,295	23,618	324	glycoside hydrolase family 4	COG1486 Alpha-galactosidases/6-phospho-beta-glucosidases, family 4 of glycosyl hydrolases
IYC_03786	23,714	25,066	1,353	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_03791	25,235	25,816	582	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_03796	25,903	26,946	1,044	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_03801	26,967	27,956	990	hypothetical protein	COG4632 Exopolysaccharide biosynthesis protein related to N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase
IYC_03806	27,971	29,137	1,167	hypothetical protein	
IYC_03811	29,220	30,671	1,452	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_03816	30,683	31,378	696	transcriptional regulatory protein	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_03821	31,463	32,059	597	hypothetical protein	
IYC_03826	32,432	33,082	651	hemolysin III	COG1272 Predicted membrane protein, hemolysin III homolog
Genomic Scaffold: NZ_JH470492					
IYC_03833	<1	>201	>201	D-proline reductase, PrdA proprotein	
Genomic Scaffold: NZ_JH470493					
IYC_03838	264	785	522	hypothetical protein	
IYC_03843	863	1,849	987	putative transporter	COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily
IYC_03848	2,188	2,556	369	hypothetical protein	COG2832 Uncharacterized protein conserved in bacteria
IYC_03853	2,562	4,367	1,806	ABC transporter, permease/ATP-binding protein	COG4988 ABC-type transport system involved in cytochrome bd biosynthesis, ATPase and permease components
IYC_03858	4,369	6,114	1,746	ABC transporter, ATP-binding/permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_03863	6,169	6,483	315	hypothetical protein	
IYC_03868	6,698	7,384	687	Crp/Fnr family transcriptional regulator	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_03873	7,501	8,700	1,200	major facilitator superfamily protein	COG0477 Permeases of the major facilitator superfamily
IYC_03878	8,999	9,913	915	iron chelate ABC transporter solute-binding protein	COG0614 ABC-type Fe3+-hydroxamate transport system,periplasmic component
IYC_03883	9,935	10,933	999	iron chelate ABC transporter permease protein	COG0609 ABC-type Fe3+-siderophore transport system,permease component
IYC_03888	10,936	11,697	762	iron chelate ABC transporter ATP-binding protein	COG1120 ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components
IYC_03893	12,194	13,171	978	lipoyltransferase and lipoate-protein ligase	COG0095 Lipoate-protein ligase A
IYC_03898	13,220	14,068	849	lipoyl synthase	COG0320 Lipoate synthase
IYC_03903	14,311	14,634	324	hypothetical protein	
IYC_03908	14,707	15,984	1,278	iron-sulfur cluster-binding protein	COG1145 Ferredoxin
IYC_03913	16,384	17,229	846	RpiR family transcriptional regulator	COG1737 Transcriptional regulators
IYC_03918	17,509	18,417	909	N-acetylmuramic acid-6-phosphate etherase	COG2103 Predicted sugar phosphate isomerase
IYC_03923	18,542	19,897	1,356	PTS system, IIBC components	COG1264 Phosphotransferase system IIB components
IYC_03928	19,961	21,037	1,077	outer surface protein	COG3589 Uncharacterized conserved protein
IYC_03933	21,344	22,195	852	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_03938	22,240	22,659	420	hypothetical protein	
IYC_03943	22,659	23,579	921	radical SAM-superfamily protein	COG1533 DNA repair photolyase
IYC_03948	23,910	24,362	453	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_03953	24,456	25,610	1,155	putative major facilitator superfamily protein	COG0477 Permeases of the major facilitator superfamily
IYC_03958	25,678	27,039	1,362	Heme biosynthesis (nirJ-2) family protein	COG0641 Arylsulfatase regulator (Fe-S oxidoreductase)
IYC_03963	27,240	27,611	372	hypothetical protein	
IYC_03968	27,787	28,395	609	hypothetical protein	COG3554 Uncharacterized protein conserved in bacteria
IYC_03973	28,486	29,127	642	hypothetical protein	COG1139 Uncharacterized conserved protein containing a ferredoxin-like domain
IYC_03978	29,366	31,237	1,872	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_03983	31,287	32,507	1,221	putative cell-surface protein	
IYC_03988	32,507	33,049	543	hypothetical protein	
IYC_03993	33,224	33,493	270	hypothetical protein	
IYC_03998	33,853	35,037	1,185	peptidase	COG1473 Metal-dependent amidase/aminoacylase/carboxypeptidase
IYC_04003	35,325	36,116	792	amino acid ABC transporter periplasmic protein	COG0834 ABC-type amino acid transport/signal transduction systems, periplasmic component/domain
IYC_04008	36,201	36,848	648	amino acid ABC transporter permease	COG0765 ABC-type amino acid transport system, permease component
IYC_04013	36,870	37,604	735	polar amino acid ABC transporter ATP-binding protein	COG1126 ABC-type polar amino acid transport system,ATPase component
IYC_04018	37,806	39,524	1,719	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_04023	39,732	41,447	1,716	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_04028	41,734	43,323	1,590	peptide/opine/nickel uptake ABC transporter substrate-binding protein	COG0747 ABC-type dipeptide transport system, periplasmic component
IYC_04033	43,320	44,276	957	peptide/opine/nickel uptake ABC transporter permease	COG0601 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components
IYC_04038	44,285	45,085	801	peptide/opine/nickel uptake ABC transporter permease	COG1173 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components
IYC_04043	45,184	45,978	795	peptide/opine/nickel uptake ABC transporter ATP-binding protein	COG0444 ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component
IYC_04048	45,972	46,784	813	peptide/opine/nickel uptake ABC transporter ATP-binding protein	COG1123 ATPase components of various ABC-type transport systems, contain duplicated ATPase
IYC_04053	47,248	47,763	516	putative lipoprotein	
IYC_04058	48,052	48,840	789	manganese/zinc/iron chelate uptake ABC transporter permease	COG1108 ABC-type Mn2+/Zn2+ transport systems, permease components
IYC_04063	48,840	49,493	654	manganese/zinc/iron chelate ABC transporter ATP-binding protein	COG1121 ABC-type Mn/Zn transport systems, ATPase component
IYC_04068	49,714	50,643	930	manganese/zinc/iron chelate ABC transporter substrate-binding protein	COG0803 ABC-type metal ion transport system, periplasmic component/surface adhesin
IYC_04073	50,767	51,261	495	hypothetical protein	
IYC_04078	51,650	53,416	1,767	helicase domain-containing protein	COG0513 Superfamily II DNA and RNA helicases
IYC_04083	53,553	54,101	549	GGDEF domain-containing protein	COG2199 FOG: GGDEF domain
IYC_04088	54,201	54,308	108	hypothetical protein	
IYC_04093	54,534	55,463	930	hypothetical protein	COG1840 ABC-type Fe3+ transport system, periplasmiccomponent
IYC_04098	55,438	56,214	777	ABC transporter-like protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_04103	56,376	57,068	693	cobalamin synthesis protein p47k	COG0378 Ni2+-binding GTPase involved in regulation of expression and maturation of urease and hydrogenase
IYC_04108	57,214	57,945	732	CRP/FNR family transcriptional regulator	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_04113	58,074	59,081	1,008	putative lipoprotein	COG0715 ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components
IYC_04118	59,053	59,814	762	putative ABC transporter permease	COG0600 ABC-type nitrate/sulfonate/bicarbonate transport system, permease component
IYC_04123	59,968	60,567	600	ABC transporter ATP-binding protein	COG1116 ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component
IYC_04128	60,747	63,257	2,511	ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_04133	63,435	64,124	690	ABC transporter ATPase	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_04138	64,251	64,406	156	hypothetical protein	COG1433 Uncharacterized conserved protein
IYC_04143	64,381	64,854	474	Fur family transcriptional regulator	COG0735 Fe2+/Zn2+ uptake regulation proteins
IYC_04148	64,824	65,297	474	hypothetical protein	COG1342 Predicted DNA-binding proteins
IYC_04153	65,405	66,676	1,272	zinc carboxypeptidase family protein	COG3409 Putative peptidoglycan-binding domain-containing protein
IYC_04158	67,048	68,088	1,041	ribonuclease Z	COG1234 Metal-dependent hydrolases of the beta-lactamase superfamily III
IYC_04163	68,718	69,095	378	hypothetical protein	
IYC_04168	69,107	70,264	1,158	sensor histidine kinase	COG4585 Signal transduction histidine kinase
IYC_04173	70,317	70,970	654	LuxR family DNA-binding response regulator	COG2197 Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain
IYC_04178	71,227	72,165	939	ABC transporter, ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_04183	72,274	73,455	1,182	putative ABC transporter, permease protein	COG0842 ABC-type multidrug transport system, permease component
IYC_04188	73,457	74,554	1,098	hypothetical protein	COG0842 ABC-type multidrug transport system, permease component
IYC_04193	74,573	74,791	219	hypothetical protein	
IYC_04198	75,325	75,876	552	nitroreductase family protein	COG0778 Nitroreductase
IYC_04203	76,411	76,902	492	hypothetical protein	
IYC_04208	77,070	77,273	204	major cold shock protein	COG1278 Cold shock proteins
IYC_04213	77,754	79,217	1,464	aminoacyl-histidine dipeptidase	COG2195 Di- and tripeptidases
IYC_04218	79,277	80,170	894	TPR domain-containing protein	COG0457 FOG: TPR repeat
IYC_04223	80,422	81,258	837	pyridoxamine kinase	COG2240 Pyridoxal/pyridoxine/pyridoxamine kinase
IYC_04228	81,555	82,193	639	putative methionine-S-sulfoxide reductase	COG0225 Peptide methionine sulfoxide reductase
IYC_04233	82,566	82,658	93	hypothetical protein	
IYC_04238	82,747	83,511	765	peptide ABC transporter ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_04243	83,498	85,387	1,890	peptide ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_04248	85,536	86,033	498	hypothetical protein	COG0741 Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)
IYC_04253	86,494	87,195	702	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_04258	87,197	88,174	978	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_04263	88,230	88,985	756	peptide ABC transporter ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_04268	88,975	90,783	1,809	peptide ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_04273	90,852	91,574	723	hypothetical protein	
IYC_04278	91,694	92,728	1,035	hypothetical protein	
IYC_04283	93,122	94,300	1,179	class I/II aminotransferase	COG1168 Bifunctional PLP-dependent enzyme with beta-cystathionase and maltose regulon repressor activities
IYC_04288	94,487	95,305	819	amino acid ABC transporter amino acid-binding protein	COG0834 ABC-type amino acid transport/signal transduction systems, periplasmic component/domain
IYC_04293	95,318	95,983	666	amino acid ABC transporter permease	COG0765 ABC-type amino acid transport system, permease component
IYC_04298	96,002	96,739	738	phosphate ABC transporter ATP-binding protein	COG1126 ABC-type polar amino acid transport system,ATPase component
IYC_04303	96,956	97,174	219	putative DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_04308	97,375	98,649	1,275	D-alanyl-D-alanine carboxypeptidase family protein	COG1686 D-alanyl-D-alanine carboxypeptidase
IYC_04313	98,972	99,595	624	MutT/NUDIX NTP pyrophosphatase	COG0494 NTP pyrophosphohydrolases including oxidative damage repair enzymes
IYC_04318	99,646	99,993	348	hypothetical protein	COG0477 Permeases of the major facilitator superfamily

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_04323	100,363	101,526	1,164	NADH-dependent butanol dehydrogenase	COG1979 Uncharacterized oxidoreductases, Fe-dependent alcohol dehydrogenase family
IYC_04328	101,853	102,077	225	hypothetical protein	
IYC_04333	102,209	103,720	1,512	C4-dicarboxylate transporter	COG1288 Predicted membrane protein
IYC_04338	103,971	105,452	1,482	nicotinate phosphoribosyltransferase	COG1488 Nicotinic acid phosphoribosyltransferase
IYC_04343	105,911	106,720	810	polar amino acid ABC transporter amino acid-binding protein	COG0834 ABC-type amino acid transport/signal transduction systems, periplasmic component/domain
IYC_04348	106,722	107,381	660	His/Glu/Gln/Arg/opine ABC transporter permease	COG0765 ABC-type amino acid transport system, permease component
IYC_04353	107,510	108,058	549	hypothetical protein	COG5418 Predicted secreted protein
IYC_04358	108,153	110,165	2,013	hypothetical protein	COG0210 Superfamily I DNA and RNA helicases
IYC_04363	110,506	110,808	303	hypothetical protein	
IYC_04368	111,013	113,256	2,244	putative cell division protein FtsA	COG0849 Actin-like ATPase involved in cell division
IYC_04373	113,344	114,522	1,179	amidohydrolase family protein	COG1473 Metal-dependent amidase/aminoacylase/carboxypeptidase
IYC_04378	114,577	114,798	222	N-acetylmuramoyl-L-alanine amidase	COG0860 N-acetylmuramoyl-L-alanine amidase
IYC_04383	114,908	115,111	204	hypothetical protein	
IYC_04388	115,746	116,699	954	RluA family pseudouridine synthase	COG0564 Pseudouridylate synthases, 23S RNA-specific
IYC_04393	116,732	118,339	1,608	stage V sporulation protein B	COG2244 Membrane protein involved in the export of O-antigen and teichoic acid
IYC_04398	118,375	119,880	1,506	stage V sporulation protein B	COG2244 Membrane protein involved in the export of O-antigen and teichoic acid
IYC_04403	120,128	120,691	564	hypothetical protein	
IYC_04408	120,712	121,119	408	hypothetical protein	COG0311 Predicted glutamine amidotransferase involved in pyridoxine biosynthesis
IYC_04413	122,238	124,556	2,319	ATP-dependent protease	COG1067 Predicted ATP-dependent protease
IYC_04433	126,705	128,480	1,776	thermolysin metallopeptidase	COG3227 Zinc metalloprotease (elastase)
IYC_04438	128,597	130,339	1,743	bacillolysin	COG3227 Zinc metalloprotease (elastase)
IYC_04443	130,463	131,905	1,443	thermolysin metallopeptidase	COG3227 Zinc metalloprotease (elastase)
IYC_04448	132,753	133,589	837	nicotinate-nucleotide pyrophosphorylase	COG0157 Nicotinate-nucleotide pyrophosphorylase
IYC_04453	133,570	134,871	1,302	L-aspartate oxidase	COG0029 Aspartate oxidase
IYC_04458	134,874	135,788	915	quinolinate synthetase	COG0379 Quinolinate synthase
IYC_04463	135,960	136,886	927	hypothetical protein	
IYC_04468	136,973	137,611	639	hemolysin III	COG1272 Predicted membrane protein, hemolysin III homolog
IYC_04473	137,713	138,363	651	transcriptional regulator	COG1309 Transcriptional regulator
IYC_04478	138,642	139,970	1,329	putative glycine/sarcosine/betaine reductase component B subunit alpha	
IYC_04483	139,986	141,032	1,047	betaine reductase complex component B subunit gamma	
IYC_04488	141,060	141,299	240	betaine reductase complex component B subunit gamma	
IYC_04493	141,405	142,895	1,491	glycine betaine transporter	COG1292 Choline-glycine betaine transporter
IYC_04498	142,993	144,090	1,098	GTP-binding protein YchF	COG0012 Predicted GTPase, probable translation factor
IYC_04503	144,836	145,264	429	cell division protein MraZ	COG2001 Uncharacterized protein conserved in bacteria
IYC_04508	145,344	146,273	930	16S rRNA m(4)C1402 methyltransferase	COG0275 Predicted S-adenosylmethionine-dependent methyltransferase involved in cell envelope biogenesis
IYC_04513	146,356	146,859	504	putative cell division protein FtsL	COG4839 Protein required for the initiation of celldivision
IYC_04518	146,875	149,055	2,181	stage V sporulation protein D	COG0768 Cell division protein FtsI/penicillin-binding protein 2
IYC_04523	149,300	150,751	1,452	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase	COG0769 UDP-N-acetylmuramyl tripeptide synthase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_04528	150,865	152,253	1,389	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D- alanine ligase	COG0770 UDP-N-acetylmuramyl pentapeptide synthase
IYC_04533	152,289	153,242	954	phospho-N-acetylmuramoyl-pentapeptide- transferase	COG0472 UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase
IYC_04538	153,321	154,433	1,113	cell cycle protein FtsW	COG0772 Bacterial cell division membrane protein
IYC_04543	154,755	155,525	771	cell division protein FtsQ	COG1589 Cell division septal protein
IYC_04548	155,536	156,249	714	hypothetical protein	COG3879 Uncharacterized protein conserved in bacteria
IYC_04553	156,282	156,632	351	small basic protein	COG3856 Uncharacterized conserved protein (small basic protein)
IYC_04558	156,633	157,364	732	hypothetical protein	COG3879 Uncharacterized protein conserved in bacteria
IYC_04563	157,457	158,128	672	hypothetical protein	COG0325 Predicted enzyme with a TIM-barrel fold
IYC_04568	158,149	158,601	453	cell division protein sepF	COG1799 Uncharacterized protein conserved in bacteria
IYC_04573	158,617	158,892	276	YGGT family protein	COG0762 Predicted integral membrane protein
IYC_04578	158,893	159,663	771	hypothetical protein	COG2302 Uncharacterized conserved protein, containsS4-like domain
IYC_04583	159,679	160,338	660	cell-division initiation protein DivIVA	COG3599 Cell division initiation protein
IYC_04588	160,646	161,734	1,089	3-dehydroquinate synthase	COG0337 3-dehydroquinate synthetase
IYC_04593	161,735	162,427	693	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	COG0775 Nucleoside phosphorylase
IYC_04598	162,548	163,348	801	3-deoxy-7-phosphoheptulonate synthase	COG2876 3-deoxy-D-arabino-heptulosonate 7-phosphate(DAHP) synthase
IYC_04603	163,515	164,843	1,329	3-phosphoshikimate 1-carboxyvinyltransferase	COG0128 5-enolpyruvylshikimate-3-phosphate synthase
IYC_04608	165,061	165,684	624	putative RNA polymerase-binding protein DksA	COG1734 DnaK suppressor protein
IYC_04613	165,786	165,905	120	hypothetical protein	
IYC_04618	166,005	166,448	444	lipoprotein signal peptidase	COG0597 Lipoprotein signal peptidase
IYC_04623	166,452	167,369	918	RluA family pseudouridine synthase	COG0564 Pseudouridylate synthases, 23S RNA-specific
IYC_04628	167,834	168,370	537	bifunctional pyrimidine regulatory protein PyrR uracil phosphoribosyltransferase	COG2065 Pyrimidine operon attenuation protein/uracil phosphoribosyltransferase
IYC_04633	168,414	169,757	1,344	uracil permease	COG2233 Xanthine/uracil permeases
IYC_04638	170,304	171,065	762	exodeoxyribonuclease III	COG0708 Exonuclease III
IYC_04643	171,287	172,429	1,143	putative methyltransferase	COG0116 Predicted N6-adenine-specific DNA methylase
IYC_04648	172,783	173,328	546	hypothetical protein	COG0288 Carbonic anhydrase
IYC_04653	173,507	174,322	816	molybdate ABC transporter, periplasmic molybdate-binding protein	COG0725 ABC-type molybdate transport system, periplasmic component
IYC_04658	174,332	175,003	672	molybdate ABC transporter, permease protein	COG4149 ABC-type molybdate transport system, permease component
IYC_04663	175,201	175,851	651	putative molybdate ABC transporter ATP-binding protein	COG3842 ABC-type spermidine/putrescine transport systems, ATPase components
IYC_04668	175,918	176,490	573	hypothetical protein	
IYC_04673	176,941	177,861	921	patatin-like phospholipase family protein	COG1752 Predicted esterase of the alpha-beta hydrolase superfamily
IYC_04678	177,953	178,108	156	hypothetical protein	
IYC_04683	178,404	179,456	1,053	putative methylcobalamin:coenzyme M methyltransferase	COG0407 Uroporphyrinogen-III decarboxylase
IYC_04688	179,504	180,901	1,398	Na+/H+ antiporter family protein	COG1757 Na+/H+ antiporter
IYC_04693	181,226	182,956	1,731	iron-sulfur cluster-binding protein	COG3894 Uncharacterized metal-binding protein
IYC_04698	183,044	183,973	930	CobW/P47K family protein	COG0523 Putative GTPases (G3E family)

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_04703	183,998	184,870	873	putative methylcobalamin:coenzyme M methyltransferase	COG0407 Uroporphyrinogen-III decarboxylase
IYC_04708	184,896	185,558	663	dimethylamine corrinoid protein	COG5012 Predicted cobalamin binding protein
IYC_04713	185,785	186,807	1,023	methylcobalamin:coenzyme M methyltransferase	COG0407 Uroporphyrinogen-III decarboxylase
IYC_04718	187,653	188,723	1,071	putative methylcobalamin:coenzyme M methyltransferase	COG0407 Uroporphyrinogen-III decarboxylase
IYC_04723	188,745	189,707	963	methionine ABC transporter ATP-binding protein	COG1135 ABC-type metal ion transport system, ATPasecomponent
IYC_04728	189,709	190,365	657	methionine ABC transporter permease	COG2011 ABC-type metal ion transport system, permease component
IYC_04733	190,639	191,457	819	NLPA family lipoprotein	COG1464 ABC-type metal ion transport system, periplasmic component/surface antigen
IYC_04738	191,668	192,249	582	MORN repeat-containing protein	COG4642 Uncharacterized protein conserved in bacteria
IYC_04743	192,572	193,591	1,020	FAD/FMN-binding family oxidoreductase	COG1902 NADH:flavin oxidoreductases, Old Yellow Enzyme family
IYC_04748	193,906	194,928	1,023	asparagine synthetase AsnA	COG2502 Asparagine synthetase A
IYC_04753	195,033	196,598	1,566	oxidoreductase, FAD dependent	COG0665 Glycine/D-amino acid oxidases (deaminating)
IYC_04758	196,952	198,811	1,860	DNA topoisomerase	COG0550 Topoisomerase IA
IYC_04763	198,837	200,744	1,908	MutS domain-containing protein	COG0249 Mismatch repair ATPase (MutS family)
IYC_04768	201,061	202,197	1,137	aldo/keto reductase family oxidoreductase	COG1453 Predicted oxidoreductases of the aldo/keto reductase family
IYC_04773	202,262	202,753	492	hypothetical protein	COG3760 Uncharacterized conserved protein
IYC_04778	203,023	203,721	699	peptide ABC transporter ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_04783	203,722	205,935	2,214	efflux ABC transporter, permease protein	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_04788	205,947	206,570	624	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_04793	206,931	207,482	552	hypothetical protein	COG5418 Predicted secreted protein
IYC_04798	207,495	208,592	1,098	carbohydrate uptake ABC transporter ATP-binding protein	COG3839 ABC-type sugar transport systems, ATPase components
IYC_04803	208,585	209,478	894	carbohydrate ABC transporter permease	COG1175 ABC-type sugar transport systems, permease components
IYC_04808	209,514	210,338	825	carbohydrate ABC transporter permease	COG0395 ABC-type sugar transport system, permease component
IYC_04813	210,384	211,745	1,362	carbohydrate uptake ABC transporter carbohydrate-binding protein	COG1653 ABC-type sugar transport system, periplasmic component
IYC_04818	211,969	212,925	957	L-lactate dehydrogenase	COG0039 Malate/lactate dehydrogenases
IYC_04823	213,427	214,137	711	GntR family transcriptional regulator	COG2186 Transcriptional regulators
IYC_04828	214,574	215,062	489	redox family protein	
IYC_04833	215,296	216,237	942	C4-dicarboxylate transporter/malic acid transport protein	COG1275 Tellurite resistance protein and related permeases
IYC_04838	216,622	219,177	2,556	putative sensory box-containing diguanylate cyclase	COG2202 FOG: PAS/PAC domain
IYC_04843	219,283	219,630	348	hypothetical protein	COG1145 Ferredoxin
IYC_04848	219,925	221,997	2,073	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_04853	222,274	222,741	468	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_04858	222,768	223,310	543	hypothetical protein	COG0716 Flavodoxins
IYC_04863	223,463	223,816	354	putative transcriptional regulator	COG1733 Predicted transcriptional regulators
IYC_04868	223,946	224,068	123	putative nitroreductase	COG0778 Nitroreductase
IYC_04873	224,219	224,584	366	glyoxalase family protein	COG0346 Lactoylglutathione lyase and related lyases

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_04878	224,613	225,083	471	methylated-DNA--protein-cysteine methyltransferase	COG0350 Methylated DNA-protein cysteine methyltransferase
IYC_04883	225,134	226,321	1,188	AraC family transcriptional regulator	COG5564 Predicted TIM-barrel enzyme, possibly a dioxygenase
IYC_04888	226,473	227,681	1,209	hypothetical protein	COG5441 Uncharacterized conserved protein
IYC_04893	227,706	228,536	831	hypothetical protein	COG5564 Predicted TIM-barrel enzyme, possibly a dioxygenase
IYC_04898	228,557	228,994	438	putative cupin	COG2140 Thermophilic glucose-6-phosphate isomerase and related metalloenzymes
IYC_04903	229,379	230,485	1,107	putative DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_04908	230,640	231,362	723	membrane protein, YKJA-BACSU-like protein	COG2323 Predicted membrane protein
IYC_04913	231,560	231,928	369	hypothetical protein	
IYC_04918	231,940	232,632	693	hypothetical protein	COG2323 Predicted membrane protein
IYC_04923	232,691	232,900	210	hypothetical protein	
IYC_04928	232,924	233,604	681	manganese catalase	COG3546 Mn-containing catalase
IYC_04933	233,891	234,046	156	hypothetical protein	
IYC_04938	234,277	234,639	363	putative iron-dependent repressor	COG1321 Mn-dependent transcriptional regulator
IYC_04943	234,696	234,866	171	putative multidrug export ATP-binding/permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_04948	234,940	235,860	921	hypothetical protein	COG5504 Predicted Zn-dependent protease
IYC_04953	236,047	236,415	369	hypothetical protein	
IYC_04958	236,789	237,640	852	GCN5-related N-acetyltransferase	COG0456 Acetyltransferases
IYC_04963	237,737	238,999	1,263	l-lysine 2,3-aminomutase	COG1509 Lysine 2,3-aminomutase
IYC_04968	239,240	240,094	855	HDIG domain protein	COG1639 Predicted signal transduction protein
IYC_04973	240,334	241,188	855	DNA repair exonuclease family protein	COG1408 Predicted phosphohydrolases
IYC_04978	241,384	244,020	2,637	cation-transporting ATPase, P-type	COG0474 Cation transport ATPase
IYC_04983	244,842	245,516	675	two component transcriptional regulator, winged helix family protein	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_04988	245,513	246,547	1,035	sensory transduction histidine kinase	COG0642 Signal transduction histidine kinase
IYC_04993	246,667	248,205	1,539	beta-lactamase	COG1680 Beta-lactamase class C and other penicillinbinding proteins
IYC_04998	248,317	249,084	768	ABC transporter related protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_05003	249,071	250,975	1,905	hypothetical protein	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_05008	251,261	251,701	441	hypothetical protein	
IYC_05013	251,715	252,140	426	hypothetical protein	
IYC_05018	252,256	252,951	696	cyclic nucleotide-binding domain-containing protein	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_05023	253,073	254,011	939	ABC-type MDR transport system, ATPase	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_05028	254,022	254,765	744	ABC-type MDR transport system, permease	COG0842 ABC-type multidrug transport system, permease component
IYC_05033	254,765	255,754	990	ABC-type MDR transport system, permease	COG0842 ABC-type multidrug transport system, permease component
IYC_05048	258,645	259,367	723	transcription activator effector binding domain-containing protein	COG3708 Uncharacterized protein conserved in bacteria
IYC_05053	259,425	260,369	945	hypothetical protein	COG2378 Predicted transcriptional regulator
IYC_05058	260,828	262,249	1,422	stage V sporulation protein AF	
IYC_05063	262,610	263,203	594	regulatory protein, tetr	COG1309 Transcriptional regulator
IYC_05068	263,235	264,269	1,035	amidohydrolase 2	COG2159 Predicted metal-dependent hydrolase of the TIM-barrel fold
IYC_05073	264,661	265,656	996	putative 2-hydroxyglutaryl-CoA dehydratase subunit	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB
IYC_05078	265,834	266,595	762	putative CoA-substrate-specific enzyme activase	COG1924 Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)
IYC_05083	266,612	267,043	432	queuosine biosynthesis protein QueD	COG0720 6-pyruvoyl-tetrahydropterin synthase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_05088	267,045	267,710	666	radical SAM domain-containing protein	COG0602 Organic radical activating enzymes
IYC_05093	267,714	268,304	591	GTP cyclohydrolase I	COG0302 GTP cyclohydrolase I
IYC_05098	268,475	269,134	660	exsB protein	COG0603 Predicted PP-loop superfamily ATPase
IYC_05103	269,261	270,145	885	AraC family transcription regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_05108	270,190	270,657	468	hypothetical protein	
IYC_05113	270,937	271,755	819	caax amino protease family protein	COG1266 Predicted metal-dependent membrane protease
IYC_05118	271,882	272,838	957	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_05123	273,005	273,757	753	hypothetical protein	COG5523 Predicted integral membrane protein
IYC_05128	274,135	274,773	639	deoxyribose-phosphate aldolase	COG0274 Deoxyribose-phosphate aldolase
IYC_05133	275,141	276,025	885	ribokinase	COG0524 Sugar kinases, ribokinase family
IYC_05138	276,288	277,559	1,272	ATP phosphoribosyltransferase regulatory subunit	COG3705 ATP phosphoribosyltransferase involved in histidine biosynthesis
IYC_05143	277,580	278,218	639	ATP phosphoribosyltransferase catalytic subunit	COG0040 ATP phosphoribosyltransferase
IYC_05148	278,295	279,599	1,305	bifunctional histidinal dehydrogenase/ histidinol dehydrogenase	COG0141 Histidinol dehydrogenase
IYC_05153	279,749	280,339	591	imidazoleglycerol-phosphate dehydratase	COG0131 Imidazoleglycerol-phosphate dehydratase
IYC_05158	280,468	281,073	606	imidazole glycerol phosphate synthase subunit HisH	COG0118 Glutamine amidotransferase
IYC_05163	281,070	281,798	729	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase	COG0106 Phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide (ProFAR) isomerase
IYC_05168	281,795	282,859	1,065	histidinol-phosphate aminotransferase	COG0079 Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase
IYC_05173	283,071	283,832	762	imidazole glycerol phosphate synthase subunit HisF	COG0107 Imidazoleglycerol-phosphate synthase
IYC_05178	283,987	284,319	333	phosphoribosyl-AMP cyclohydrolase	COG0139 Phosphoribosyl-AMP cyclohydrolase
IYC_05183	284,445	284,777	333	phosphoribosyl-ATP pyrophosphatase	COG0140 Phosphoribosyl-ATP pyrophosphohydrolase
IYC_05188	284,915	285,610	696	hypothetical protein	COG2391 Predicted transporter component
IYC_05193	285,624	285,845	222	hypothetical protein	COG0425 Predicted redox protein, regulator of disulfide bond formation
IYC_05198	285,863	286,390	528	hypothetical protein	COG2391 Predicted transporter component
IYC_05203	286,421	287,311	891	LysR family transcriptional regulator	COG0583 Transcriptional regulator
IYC_05208	287,577	288,827	1,251	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter
IYC_05213	288,910	290,610	1,701	pyridine nucleotide-disulfide oxidoreductase family protein	COG0446 Uncharacterized NAD(FAD)-dependent dehydrogenases
IYC_05218	290,763	292,019	1,257	hypothetical protein	COG2391 Predicted transporter component
IYC_05223	292,303	293,010	708	antibiotic ABC transporter ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_05228	293,013	293,744	732	antibiotic ABC transporter, permease protein	
IYC_05233	293,747	294,490	744	putative lantibiotic ABC transporter permease	
IYC_05238	294,639	295,298	660	putative lantibiotic biosynthesis regulatory protein	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_05243	295,289	296,665	1,377	putative antibiotic transporter sensor protein	COG0642 Signal transduction histidine kinase
IYC_05248	297,232	298,458	1,227	arginine deiminase	COG2235 Arginine deiminase
IYC_05253	298,583	300,001	1,419	arginine/ornithine antiporter	COG0531 Amino acid transporters
Genomic Scaffold: NZ_JH470494					

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_05258	314	955	642	RNA polymerase factor sigma-70	COG1595 DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
IYC_05263	1,029	1,538	510	hypothetical protein	COG3688 Predicted RNA-binding protein containing a PIN domain
IYC_05268	1,541	2,314	774	TrmH family RNA methyltransferase	COG0566 rRNA methylases
IYC_05273	2,317	3,069	753	FAD-dependent thymidylate synthase	COG1351 Predicted alternative thymidylate synthase
IYC_05278	3,059	3,499	441	RNase3 domain-containing protein	COG1939 Uncharacterized protein conserved in bacteria
IYC_05283	3,548	4,945	1,398	cysteinyl-tRNA synthetase	COG0215 Cysteinyl-tRNA synthetase
IYC_05288	4,974	6,191	1,218	threonine dehydratase	COG1171 Threonine dehydratase
IYC_05293	6,263	7,900	1,638	putative prolyl-tRNA synthetase	COG0442 Prolyl-tRNA synthetase
IYC_05298	8,006	8,695	690	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	COG1211 4-diphosphocytidyl-2-methyl-D-erithritol synthase
IYC_05303	8,697	9,797	1,101	PIN/TRAM domain-containing protein	COG4956 Integral membrane protein (PIN domain superfamily)
IYC_05308	10,198	10,596	399	hypothetical protein	
IYC_05313	10,628	11,689	1,062	DNA integrity scanning protein DisA	COG1623 Predicted nucleic-acid-binding protein (contains the HHH domain)
IYC_05318	11,708	13,066	1,359	DNA repair protein RadA	COG1066 Predicted ATP-dependent serine protease
IYC_05323	13,186	15,621	2,436	negative regulator of genetic competence MecB/ClpC	COG0542 ATPases with chaperone activity, ATP-binding subunit
IYC_05328	15,623	16,666	1,044	ATP:guanido phosphotransferase	COG3869 Arginine kinase
IYC_05333	16,671	17,198	528	UVR domain-containing protein	COG3880 Uncharacterized protein with conserved CXXCpairs
IYC_05338	17,221	17,685	465	transcriptional regulator CtsR	COG4463 Transcriptional repressor of class III stress genes
IYC_05343	18,033	18,263	231	hypothetical protein	
IYC_05348	18,630	20,693	2,064	elongation factor G	COG0480 Translation elongation factors (GTPases)
Genomic Scaffold: NZ_JH470495					
IYC_05359	<1	2,124	>2124	phage protein	COG5283 Phage-related tail protein
IYC_05374	2,925	3,026	102	putative phage tail protein	
IYC_05379	3,026	>3489	>464	phage endopeptidase	COG4926 Phage-related protein
Genomic Scaffold: NZ_JH470496					
IYC_05384	<1	>282	>282	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
Genomic Scaffold: NZ_JH470497					
IYC_05399	437	712	276	hypothetical protein	
IYC_05414	1,860	2,636	777	PrdD protein	
IYC_05419	2,650	3,120	471	PrdE protein	
IYC_05424	3,134	4,141	1,008	proline racemase	COG3938 Proline racemase
IYC_05429	4,629	5,498	870	hypothetical protein	
IYC_05434	5,647	7,092	1,446	amino acid permease family protein	COG0531 Amino acid transporters
Genomic Scaffold: NZ_JH470498					
IYC_05449	2,395	2,607	213	hypothetical protein	
IYC_05454	2,708	3,154	447	D-proline reductase subunit gamma PrdB	
IYC_05459	3,182	3,427	246	D-proline reductase, PrdB subunit	
IYC_05464	3,795	4,673	879	hypothetical protein	
IYC_05469	5,008	6,006	999	phosphotransacetylase	COG0280 Phosphotransacetylase
IYC_05474	6,145	7,338	1,194	acetate kinase A/propionate kinase 2	COG0282 Acetate kinase
IYC_05479	7,456	7,956	501	ribosomal protein L32 family protein	COG1399 Predicted metal-binding, possibly nucleic acid-binding protein
IYC_05484	7,970	8,152	183	50S ribosomal protein L32	COG0333 Ribosomal protein L32
IYC_05489	8,242	9,249	1,008	putative phosphate acyltransferase	COG0416 Fatty acid/phospholipid biosynthesis enzyme
IYC_05494	9,327	9,560	234	acyl carrier protein	COG0236 Acyl carrier protein
IYC_05499	9,685	10,389	705	ribonuclease III	COG0571 dsRNA-specific ribonuclease
IYC_05504	10,382	11,473	1,092	radical SAM domain-containing protein	COG1243 Histone acetyltransferase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_05509	11,626	11,724	99	hypothetical protein	
IYC_05514	11,737	15,318	3,582	chromosome segregation protein SMC	COG1196 Chromosome segregation ATPases
IYC_05519	15,348	16,259	912	signal recognition particle-docking protein FtsY	COG0552 Signal recognition particle GTPase
IYC_05524	16,487	16,819	333	putative DNA-binding protein	COG2739 Uncharacterized protein conserved in bacteria
IYC_05529	16,832	18,181	1,350	signal recognition particle protein	COG0541 Signal recognition particle GTPase
IYC_05534	18,212	18,460	249	30S ribosomal protein S16	COG0228 Ribosomal protein S16
IYC_05539	18,481	18,708	228	hypothetical protein	COG1837 Predicted RNA-binding protein (contains KH domain)
IYC_05544	18,822	19,316	495	16S rRNA-processing protein RimM	COG0806 RimM protein, required for 16S rRNA processing
IYC_05549	19,307	20,029	723	tRNA (guanine-N(1)-)-methyltransferase	COG0336 tRNA-(guanine-N1)-methyltransferase
IYC_05554	20,181	20,525	345	50S ribosomal protein L19	COG0335 Ribosomal protein L19
IYC_05559	20,584	21,108	525	signal peptidase I	COG0681 Signal peptidase I
IYC_05564	21,406	22,305	900	GTPase YlqF	COG1161 Predicted GTPases
IYC_05569	22,417	23,226	810	ribonuclease HII	COG0164 Ribonuclease HII
IYC_05574	23,486	23,857	372	hypothetical protein	COG0792 Predicted endonuclease distantly related toarchaeal Holliday junction resolvase
IYC_05579	23,873	25,396	1,524	Mg chelatase-like protein	COG0606 Predicted ATPase with chaperone activity
IYC_05584	25,595	26,680	1,086	DNA protecting protein DprA	COG0758 Predicted Rossmann fold nucleotide-binding protein involved in DNA uptake
IYC_05589	27,184	29,286	2,103	DNA topoisomerase I	COG0550 Topoisomerase IA
IYC_05594	29,492	30,268	777	transcriptional repressor CodY	COG4465 Pleiotropic transcriptional repressor
IYC_05599	30,505	31,206	702	30S ribosomal protein S2	COG0052 Ribosomal protein S2
IYC_05604	31,294	32,217	924	elongation factor Ts	COG0264 Translation elongation factor Ts
IYC_05609	32,603	33,319	717	uridylate kinase	COG0528 Uridylate kinase
IYC_05614	33,335	33,889	555	ribosome recycling factor	COG0233 Ribosome recycling factor
IYC_05619	33,957	34,715	759	undecaprenyl pyrophosphate synthase	COG0020 Undecaprenyl pyrophosphate synthase
IYC_05624	34,740	35,537	798	phosphatidate cytidylyltransferase	COG0575 CDP-diglyceride synthetase
IYC_05629	35,867	36,847	981	hypothetical protein	COG0628 Predicted permease
IYC_05634	36,877	38,034	1,158	1-deoxy-D-xylulose 5-phosphate reductoisomerase	COG0743 1-deoxy-D-xylulose 5-phosphate reductoisomerase
IYC_05639	38,079	39,089	1,011	RIP metalloprotease RseP	COG0750 Predicted membrane-associated Zn-dependent proteases 1
IYC_05644	39,231	40,280	1,050	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	COG0821 Enzyme involved in the deoxyxylulose pathway of isoprenoid biosynthesis
IYC_05649	40,559	44,857	4,299	DNA polymerase III PolC	COG2176 DNA polymerase III, alpha subunit (gram-positive type)
IYC_05654	45,162	45,623	462	ribosome maturation protein RimP	COG0779 Uncharacterized protein conserved in bacteria
IYC_05659	45,640	47,013	1,374	transcription elongation factor NusA	COG0195 Transcription elongation factor
IYC_05664	47,073	47,345	273	hypothetical protein	COG2740 Predicted nucleic-acid-binding protein implicated in transcription termination
IYC_05669	47,332	47,649	318	ribosomal protein L7Ae family protein	COG1358 Ribosomal protein HS6-type (S12/L30/L7a)
IYC_05674	47,667	49,733	2,067	translation initiation factor IF-2	COG0532 Translation initiation factor 2 (IF-2; GTPase)
IYC_05679	49,764	50,126	363	ribosome-binding factor A	COG0858 Ribosome-binding factor A
IYC_05684	50,116	51,078	963	DHH family protein	COG0618 Exopolyphosphatase-related proteins
IYC_05689	51,079	51,948	870	tRNA pseudouridine synthase B	COG0130 Pseudouridine synthase
IYC_05694	51,961	52,881	921	bifunctional riboflavin kinase/FMN adenylyltransferase	COG0196 FAD synthase
IYC_05699	53,014	53,277	264	30S ribosomal protein S15	COG0184 Ribosomal protein S15P/S13E

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_05704	53,541	55,649	2,109	polynucleotide phosphorylase/polyadenylase	COG1185 Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase)
IYC_05709	55,824	57,125	1,302	M16 family peptidase	COG0612 Predicted Zn-dependent peptidases
IYC_05714	57,202	57,474	273	PRC-barrel domain-containing protein	COG1873 Uncharacterized conserved protein
IYC_05719	57,490	58,695	1,206	aspartate kinase I	COG0527 Aspartokinases
IYC_05724	58,822	59,517	696	Clp protease	COG0740 Protease subunit of ATP-dependent Clp proteases
IYC_05729	59,747	62,023	2,277	DNA translocase FtsK/SpoIIIE	COG1674 DNA segregation ATPase FtsK/SpoIIIE and related proteins
IYC_05734	62,118	63,455	1,338	RNA modification protein	COG0621 2-methylthioadenine synthetase
IYC_05739	63,439	64,023	585	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	COG0558 Phosphatidylglycerophosphate synthase
IYC_05744	64,172	65,230	1,059	recombinase A	COG0468 RecA/RadA recombinase
IYC_05749	65,464	67,005	1,542	phosphodiesterase	COG1418 Predicted HD superfamily hydrolase
IYC_05754	67,144	67,404	261	hypothetical protein	COG2359 Uncharacterized protein conserved in bacteria
IYC_05759	67,560	68,426	867	phosphoribose diphosphate:decaprenyl-phosphate phosphoribosyltransferase	COG0382 4-hydroxybenzoate polyprenyltransferase andrelated prenyltransferases
IYC_05764	68,445	69,728	1,284	hypothetical protein	
IYC_05769	69,725	70,081	357	multidrug resistance protein, SMR family	COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily
IYC_05774	70,115	71,308	1,194	aspartate aminotransferase	COG0436 Aspartate/tyrosine/aromatic aminotransferase
IYC_05779	71,419	71,676	258	phosphocarrier protein HPr	COG1925 Phosphotransferase system, HPr-related proteins
IYC_05784	71,704	71,925	222	hypothetical protein	COG2155 Uncharacterized conserved protein
IYC_05789	72,188	73,618	1,431	adenylosuccinate lyase	COG0015 Adenylosuccinate lyase
IYC_05794	73,710	74,987	1,278	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	COG0445 NAD/FAD-utilizing enzyme apparently involved in cell division
IYC_05799	75,332	75,955	624	VanZF	COG4767 Glycopeptide antibiotics resistance protein
IYC_05804	76,165	76,557	393	ArsR family transcriptional regulator	COG0640 Predicted transcriptional regulators
IYC_05809	76,644	77,261	618	methyltransferase family protein	COG0500 SAM-dependent methyltransferases
IYC_05814	77,483	78,088	606	hypothetical protein	
IYC_05819	78,577	79,386	810	hypothetical protein	COG1484 DNA replication protein
IYC_05824	79,540	80,250	711	N-acetylmuramoyl-L-alanine amidase	COG5479 Uncharacterized protein potentially involved in peptidoglycan biosynthesis
IYC_05829	80,675	81,286	612	hypothetical protein	
IYC_05834	83,077	83,193	117	hypothetical protein	
IYC_05839	83,270	>83499	>230	hypothetical protein	
Genomic Scaffold: NZ_JH470499					
IYC_05844	71	304	234	hypothetical protein	
Genomic Scaffold: NZ_JH470500					
IYC_05849	72	>725	>654	hypothetical protein	
Genomic Scaffold: NZ_JH470501					
IYC_05854	<1	>1375	>1375	putative oxidoreductase	COG0493 NADPH-dependent glutamate synthase beta chain and related oxidoreductases
Genomic Scaffold: NZ_JH470502					
IYC_05859	<1	372	>372	hypothetical protein	
IYC_05864	473	928	456	hypothetical protein	COG2216 High-affinity K+ transport system, ATPase chain B
IYC_05869	1,203	3,365	2,163	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_05874	3,343	4,035	693	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_05879	4,481	5,707	1,227	major facilitator family transporter	COG0477 Permeases of the major facilitator superfamily
IYC_05884	6,016	6,942	927	hypothetical protein	
IYC_05889	7,107	7,964	858	HAD superfamily hydrolase	COG0561 Predicted hydrolases of the HAD superfamily

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_05894	7,976	8,743	768	DeoR family transcriptional regulator	COG1349 Transcriptional regulators of sugar metabolism
IYC_05899	9,006	9,737	732	hypothetical protein	
IYC_05904	9,852	10,007	156	hypothetical protein	
IYC_05909	10,118	10,657	540	sigma-70 family RNA polymerase sigma factor	COG1191 DNA-directed RNA polymerase specialized sigma subunit
IYC_05914	10,845	11,024	180	hypothetical protein	
IYC_05919	11,148	14,801	3,654	collagenase	COG3291 FOG: PKD repeat
IYC_05924	15,308	16,195	888	5,10-methylenetetrahydrofolate reductase	COG0685 5,10-methylenetetrahydrofolate reductase
IYC_05929	16,383	17,063	681	hypothetical protein	
IYC_05934	17,053	19,431	2,379	putative 5-methyltetrahydrofolate--homocysteine methyltransferase	COG0646 Methionine synthase I (cobalamin-dependent), methyltransferase domain
IYC_05939	19,529	19,945	417	hypothetical protein	
IYC_05944	20,262	21,014	753	hypothetical protein	
IYC_05949	20,959	21,108	150	hypothetical protein	
IYC_05954	21,132	21,236	105	hypothetical protein	
IYC_05959	21,524	22,096	573	RNA polymerase factor sigma-70	COG1595 DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
IYC_05964	22,089	23,561	1,473	hypothetical protein	
IYC_05969	23,909	24,604	696	two component transcriptional regulator, winged helix family protein	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_05974	24,639	25,967	1,329	integral membrane sensor signal transduction histidine kinase	COG0642 Signal transduction histidine kinase
IYC_05979	26,324	26,584	261	hypothetical protein	
IYC_05984	26,577	26,963	387	hypothetical protein	
IYC_05989	27,001	28,974	1,974	aco operon transcriptional regulator	COG3284 Transcriptional activator of acetoin/glycerol metabolism
IYC_05994	29,274	30,257	984	TPP-dependent acetoin dehydrogenase complex, E1 component, alpha subunit	COG1071 Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, eukaryotic type, alpha subunit
IYC_05999	30,260	31,231	972	TPP-dependent acetoin dehydrogenase complex, E1 component, beta subunit	COG0022 Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, eukaryotic type, beta subunit
IYC_06004	31,318	32,628	1,311	TPP-dependent acetoin dehydrogenase complex protein	COG0508 Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes; TPP-dependent acetoin dehydrogenase complex, E2 component, dihydrolipoyllysine-residue acetyltransferase
IYC_06009	32,641	34,032	1,392	TPP-dependent acetoin dehydrogenase complex, E3 component, dihydrolipoamide dehydrogenase	COG1249 Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes
IYC_06014	34,057	35,064	1,008	putative lipoate-protein ligase A	COG0095 Lipoate-protein ligase A
IYC_06019	35,454	36,599	1,146	glycine betaine/L-proline ABC transporter, ATP-binding protein	COG1125 ABC-type proline/glycine betaine transport systems, ATPase components
IYC_06024	36,592	38,169	1,578	glycine betaine/L-proline ABC transporter, permease/glycine betaine/L-proline-binding protein	COG1174 ABC-type proline/glycine betaine transport systems, permease component
IYC_06029	38,504	39,133	630	regulatory protein TetR	COG1309 Transcriptional regulator
IYC_06034	39,422	40,024	603	hypothetical protein	
IYC_06039	40,027	40,707	681	cobalt transport protein	COG0619 ABC-type cobalt transport system, permease component CbiQ and related transporters
IYC_06044	40,718	42,094	1,377	ABC transporter	COG1122 ABC-type cobalt transport system, ATPase component
IYC_06049	43,084	45,198	2,115	type I restriction-modification system specificity subunit	COG0286 Type I restriction-modification system methyltransferase subunit

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_06054	45,204	46,583	1,380	putative type I restriction system, specificity protein HsdS	COG0732 Restriction endonuclease S subunits
IYC_06059	46,583	49,540	2,958	hypothetical protein	COG0610 Type I site-specific restriction-modification system, R (restriction) subunit and related helicases
IYC_06064	49,635	49,763	129	hypothetical protein	
IYC_06069	50,126	51,607	1,482	threonine synthase	COG0498 Threonine synthase
IYC_06074	51,617	52,504	888	homoserine kinase	COG0083 Homoserine kinase
IYC_06079	52,732	54,051	1,320	aspartate kinase	COG0527 Aspartokinases
IYC_06084	54,404	55,681	1,278	homoserine dehydrogenase	COG0460 Homoserine dehydrogenase
IYC_06089	56,926	59,112	2,187	hypothetical protein	COG0451 Nucleoside-diphosphate-sugar epimerases
IYC_06094	59,105	59,992	888	hypothetical protein	
IYC_06099	59,992	61,821	1,830	hypothetical protein	COG4878 Uncharacterized protein conserved in bacteria
IYC_06104	61,828	63,243	1,416	hypothetical protein	COG0438 Glycosyltransferase
IYC_06109	63,243	64,712	1,470	hypothetical protein	COG4267 Predicted membrane protein
IYC_06114	64,699	66,396	1,698	spore coat protein	
IYC_06119	66,416	67,117	702	hypothetical protein	
IYC_06124	67,083	67,811	729	hypothetical protein	
IYC_06129	67,801	69,216	1,416	hypothetical protein	
IYC_06134	69,542	70,153	612	transcriptional regulator, TetR family protein	COG1309 Transcriptional regulator
IYC_06139	70,655	70,963	309	transcriptional regulator PadR family protein	COG1695 Predicted transcriptional regulators
IYC_06144	71,138	71,767	630	ABC transporter related protein	COG4619 ABC-type uncharacterized transport system, ATPase component
IYC_06149	71,845	72,540	696	hypothetical protein	COG0390 ABC-type uncharacterized transport system, permease component
IYC_06154	72,594	74,504	1,911	sensory box sigma-54 dependent transcriptional regulator	COG3284 Transcriptional activator of acetoin/glycerol metabolism
IYC_06159	74,994	76,160	1,167	NADH-dependent butanol dehydrogenase	COG1454 Alcohol dehydrogenase, class IV
IYC_06164	76,638	77,525	888	ferredoxin-NADP(+) reductase subunit alpha	COG0543 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases
Genomic Scaffold: NZ_JH470503					
IYC_06169	10	1,017	1,008	adenosine deaminase	COG1816 Adenosine deaminase
IYC_06174	1,359	3,527	2,169	putative phage infection protein	COG1511 Predicted membrane protein
IYC_06179	3,554	5,713	2,160	putative phage infection protein	COG1511 Predicted membrane protein
IYC_06184	6,155	6,457	303	hypothetical protein	
IYC_06189	7,416	9,545	2,130	ATP-dependent DNA helicase RecQ	COG0514 Superfamily II DNA helicase
IYC_06194	9,608	>9717	>110	hypothetical protein	
IYC_06199	9,990	10,421	432	rubrerythrin family protein	COG1592 Rubrerythrin
IYC_06204	10,821	12,611	1,791	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_06209	12,639	13,454	816	carboxylesterase bioH, putative	COG0596 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
IYC_06214	13,778	14,323	546	phosphodiesterase	COG0622 Predicted phosphoesterase
IYC_06219	14,961	15,344	384	response regulator	COG0784 FOG: CheY-like receiver
IYC_06224	15,459	17,933	2,475	hypothetical protein	COG0457 FOG: TPR repeat
IYC_06229	18,524	19,240	717	LytTr DNA-binding domain-containing protein	COG3279 Response regulator of the LytR/AlgR family
IYC_06234	19,244	20,524	1,281	hypothetical protein	COG2972 Predicted signal transduction protein with a C-terminal ATPase domain
IYC_06239	20,711	20,854	144	hypothetical protein	
IYC_06244	20,854	21,438	585	putative AIP processing-secretion protein	COG4512 Membrane protein putatively involved in post-translational modification of the autoinducing quorum-sensing peptide
IYC_06249	21,603	22,511	909	putative lipoprotein	COG0841 Cation/multidrug efflux pump
IYC_06254	23,259	23,729	471	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_06259	23,731	24,276	546	putative cyclase	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_06264	24,576	25,760	1,185	aspartate aminotransferase	COG0436 Aspartate/tyrosine/aromatic aminotransferase
IYC_06269	26,772	27,350	579	putative cyclase	COG1878 Predicted metal-dependent hydrolase
IYC_06274	27,510	28,034	525	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_06279	28,337	29,158	822	MerR family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_06284	29,373	30,071	699	B3/4 domain-containing protein	COG3382 Uncharacterized conserved protein
IYC_06289	30,352	31,641	1,290	aspartyl-tRNA synthetase	COG0017 Aspartyl/asparaginyl-tRNA synthetases
IYC_06294	32,187	32,744	558	hypothetical protein	COG0716 Flavodoxins
IYC_06299	32,748	33,311	564	CDP-alcohol phosphatidyltransferase family protein	COG0558 Phosphatidylglycerophosphate synthase
IYC_06304	33,700	34,314	615	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_06309	34,329	34,808	480	hypothetical protein	COG4635 Flavodoxin
IYC_06314	34,940	35,629	690	HAD family hydrolase	COG1011 Predicted hydrolase (HAD superfamily)
IYC_06319	35,626	35,772	147	hypothetical protein	
IYC_06324	35,942	36,631	690	cyclic nucleotide-binding domain-containing protein	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_06329	36,941	37,525	585	DJ-1/PfpI family protein	COG0693 Putative intracellular protease/amidase
IYC_06334	37,681	38,574	894	LysR family transcriptional regulator	COG0583 Transcriptional regulator
IYC_06339	38,630	39,421	792	sigma-70 family RNA polymerase sigma factor	COG1595 DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
IYC_06344	39,498	39,944	447	hypothetical protein	
IYC_06349	40,367	40,576	210	ferrous iron transport protein	COG1918 Fe2+ transport system protein A
IYC_06354	40,593	40,814	222	ferrous iron transport protein A	COG1918 Fe2+ transport system protein A
IYC_06359	40,850	43,006	2,157	ferrous iron transport protein B	COG0370 Fe2+ transport system protein B
IYC_06364	43,027	43,173	147	hypothetical protein	
IYC_06369	43,405	43,602	198	hypothetical protein	COG2323 Predicted membrane protein
IYC_06374	43,683	44,120	438	acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_06379	44,630	45,424	795	D-alanyl-D-alanine carboxypeptidase family protein	COG1876 D-alanyl-D-alanine carboxypeptidase
IYC_06384	45,729	45,956	228	small acid-soluble spore protein Tlp	
IYC_06389	46,179	46,865	687	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_06394	46,866	48,230	1,365	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_06399	48,592	48,726	135	hypothetical protein	
IYC_06404	48,842	48,976	135	hypothetical protein	
IYC_06409	49,030	50,289	1,260	cytochrome P450	COG2124 Cytochrome P450
IYC_06414	50,317	50,478	162	hypothetical protein	
IYC_06419	50,670	52,475	1,806	heavy metal-binding domain-containing protein	COG2608 Copper chaperone
IYC_06424	52,901	53,209	309	hypothetical protein	
IYC_06429	53,621	53,857	237	putative membrane protein	COG3462 Predicted membrane protein
IYC_06434	54,118	54,819	702	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_06439	54,819	56,225	1,407	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_06444	56,565	57,245	681	hypothetical protein	
IYC_06449	57,481	58,239	759	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_06454	58,484	59,437	954	iron chelate uptake ABC transporter permease	COG4606 ABC-type enterochelin transport system, permease component
IYC_06459	59,427	60,377	951	iron chelate uptake ABC transporter permease	COG4605 ABC-type enterochelin transport system, permease component

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_06464	60,374	61,129	756	iron chelate ABC transporter ATP-binding protein	COG4604 ABC-type enterochelin transport system, ATPase component
IYC_06469	61,203	62,147	945	iron chelate ABC transporter solute-binding protein	COG4607 ABC-type enterochelin transport system, periplasmic component
IYC_06474	62,349	63,350	1,002	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_06479	63,963	65,336	1,374	sensor histidine kinase	COG2972 Predicted signal transduction protein with a C-terminal ATPase domain
IYC_06484	65,339	66,046	708	DNA-binding response regulator	COG3279 Response regulator of the LytR/AlgR family
IYC_06489	66,134	67,015	882	hypothetical protein	
IYC_06494	67,406	67,912	507	hypothetical protein	COG0452 Phosphopantothenoylcysteine synthetase/decarboxylase
IYC_06499	67,914	68,240	327	hypothetical protein	
IYC_06504	68,644	69,516	873	hypothetical protein	
IYC_06509	69,630	70,076	447	hypothetical protein	
IYC_06514	70,254	72,047	1,794	TPR repeat-containing glycosyl transferase	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_06519	72,113	>72804	>692	hypothetical protein	
Genomic Scaffold: NZ_JH470505					
IYC_06526	<1	>236	>236	hypothetical protein	
Genomic Scaffold: NZ_JH470506					
IYC_06531	<1	550	>550	hypothetical protein	
IYC_06536	655	1,275	621	hypothetical protein	
IYC_06541	1,300	1,446	147	hypothetical protein	
IYC_06546	1,515	1,763	249	hypothetical protein	COG1396 Predicted transcriptional regulators
IYC_06551	1,961	2,059	99	hypothetical protein	
IYC_06556	2,071	2,442	372	hypothetical protein	COG1396 Predicted transcriptional regulators
Genomic Scaffold: NZ_JH470507					
IYC_06561	<1	250	>250	prophage antirepressor	COG3617 Prophage antirepressor
IYC_06566	270	524	255	hypothetical protein	COG2452 Predicted site-specific integrase-resolvase
IYC_06571	517	738	222	MerR family transcriptional regulator	COG1396 Predicted transcriptional regulators
IYC_06576	963	1,283	321	post-exponential-phase responses transcriptionalregulator	COG1396 Predicted transcriptional regulators
IYC_06581	1,466	2,566	1,101	integrase family protein	COG0582 Integrase
IYC_06586	2,946	4,061	1,116	putative FAD-dependent dehydrogenase	COG0644 Dehydrogenases (flavoproteins)
IYC_06591	4,154	4,537	384	hypothetical protein	
IYC_06596	4,632	5,522	891	hypothetical protein	
IYC_06601	5,538	6,446	909	manganese/zinc/iron chelate ABC transporter substrate-binding protein	COG0803 ABC-type metal ion transport system, periplasmic component/surface adhesin
IYC_06606	6,644	7,885	1,242	HDIG domain-containing protein	COG2206 HD-GYP domain
IYC_06611	8,023	8,481	459	hypothetical protein	COG0802 Predicted ATPase or kinase
IYC_06616	8,478	9,191	714	glycoprotease family protein	COG1214 Inactive homolog of metal-dependent proteases, putative molecular chaperone
IYC_06621	9,184	9,642	459	ribosomal-protein-alanine acetyltransferase	COG0456 Acetyltransferases
IYC_06626	9,999	11,222	1,224	tyrosyl-tRNA synthetase	COG0162 Tyrosyl-tRNA synthetase
IYC_06631	11,463	11,870	408	hypothetical protein	
IYC_06636	12,007	13,164	1,158	putative ATP-binding protein	COG2865 Predicted transcriptional regulator containing an HTH domain and an uncharacterized domain shared with the mammalian protein Schlafen
IYC_06641	13,302	13,865	564	hypothetical protein	COG4720 Predicted membrane protein
IYC_06646	13,946	14,107	162	hypothetical protein	
IYC_06651	14,369	15,361	993	PAP2 family protein	COG0671 Membrane-associated phospholipid phosphatase
IYC_06656	15,461	17,050	1,590	peptide chain release factor 3	COG4108 Peptide chain release factor RF-3
IYC_06661	17,557	18,363	807	HAD family hydrolase	COG0561 Predicted hydrolases of the HAD superfamily
IYC_06666	18,697	19,695	999	tryptophanyl-tRNA synthetase	COG0180 Tryptophanyl-tRNA synthetase
IYC_06671	20,280	20,654	375	desulfoferrodoxin	COG2033 Desulfoferrodoxin
IYC_06676	20,837	21,031	195	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_06681	21,244	29,874	8,631	cyclic beta 1-2 glucan synthetase	COG3459 Cellobiose phosphorylase
IYC_06686	30,133	31,419	1,287	diaminopimelate decarboxylase	COG0019 Diaminopimelate decarboxylase
IYC_06691	31,435	32,637	1,203	aspartokinase II	COG0527 Aspartokinases
IYC_06696	33,032	34,330	1,299	putative aminopeptidase 2	COG1362 Aspartyl aminopeptidase
IYC_06701	34,480	35,658	1,179	cation efflux family protein	COG0053 Predicted Co/Zn/Cd cation transporters
IYC_06706	35,780	36,871	1,092	hypothetical protein	COG1066 Predicted ATP-dependent serine protease
IYC_06711	37,173	39,305	2,133	hypothetical protein	COG3973 Superfamily I DNA and RNA helicases
IYC_06716	39,326	41,275	1,950	ATP-dependent metalloprotease FtsH	COG0465 ATP-dependent Zn proteases
IYC_06721	41,699	43,090	1,392	dipeptidase PepV	COG0624 Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases
IYC_06726	43,293	44,729	1,437	prolyl-tRNA synthetase	COG0442 Prolyl-tRNA synthetase
IYC_06731	45,155	45,397	243	GIY-YIG domain-containing protein	COG2827 Predicted endonuclease containing a URI domain
IYC_06736	45,517	46,740	1,224	putative lipoprotein	
IYC_06741	46,841	47,710	870	ABC transporter ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_06746	47,869	48,675	807	hypothetical protein	
IYC_06751	48,650	49,222	573	RNA polymerase sigma factor	COG1595 DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
IYC_06756	49,551	51,209	1,659	sulfate permease, SulP family protein	COG0659 Sulfate permease and related transporters (MFS superfamily)
IYC_06761	51,356	52,699	1,344	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_06766	52,722	53,414	693	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_06771	53,714	54,922	1,209	putative lipoprotein	COG1277 ABC-type transport system involved in multi-copper enzyme maturation, permease component
IYC_06776	54,927	55,832	906	bacitracin ABC transporter, ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_06781	56,201	56,404	204	hypothetical protein	
IYC_06791	57,037	57,132	96	hypothetical protein	
IYC_06801	57,865	58,950	1,086	outer surface protein	COG3589 Uncharacterized conserved protein
IYC_06806	59,333	60,226	894	BadF/BadG/BcrA/BcrD ATPase family protein	COG2971 Predicted N-acetylglucosamine kinase
IYC_06811	60,286	60,582	297	PTS system lactose/cellobiose-specific transporter subunit IIA	COG1447 Phosphotransferase system cellobiose-specific component IIA
IYC_06816	60,786	61,094	309	PTS system lactose/cellobiose-specific family transporter subunit IIB	COG1440 Phosphotransferase system cellobiose-specific component IIB
IYC_06821	61,214	63,943	2,730	sigma-54 dependent transcriptional regulator	COG1221 Transcriptional regulators containing an AAA-type ATPase domain and a DNA-binding domain
IYC_06826	64,067	65,356	1,290	PTS system lactose/cellobiose family IIC subunit	COG1455 Phosphotransferase system cellobiose-specific component IIC
IYC_06831	65,439	66,776	1,338	glycosyl hydrolase, family 4	COG1486 Alpha-galactosidases/6-phospho-beta-glucosidases, family 4 of glycosyl hydrolases
IYC_06836	67,109	68,029	921	hypothetical protein	COG3595 Uncharacterized conserved protein
IYC_06841	68,019	68,120	102	hypothetical protein	
IYC_06846	68,398	68,982	585	hypothetical protein	COG1309 Transcriptional regulator
IYC_06851	69,192	69,488	297	hypothetical protein	
IYC_06856	69,968	70,477	510	ISCb1g3, transposase	COG3316 Transposase and inactivated derivatives
IYC_06871	71,513	73,201	1,689	M28 family peptidase	COG2234 Predicted aminopeptidases
IYC_06876	73,267	74,439	1,173	hypothetical protein	
IYC_06881	74,463	75,119	657	AcrR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_06886	75,339	75,620	282	hypothetical protein	
IYC_06891	75,914	76,204	291	hypothetical protein	COG4842 Uncharacterized protein conserved in bacteria
IYC_06906	77,216	77,644	429	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_06911	77,644	77,736	93	hypothetical protein	
IYC_06916	77,755	78,390	636	hypothetical protein	
IYC_06921	78,518	78,916	399	hypothetical protein	
IYC_06926	79,046	79,162	117	hypothetical protein	
IYC_06931	79,421	79,819	399	hypothetical protein	
IYC_06936	79,833	>80341	>509	YqcG	
Genomic Scaffold: NZ_JH470508					
IYC_06943	131	430	300	hypothetical protein	
IYC_06948	657	1,496	840	metallo-beta-lactamase family protein	COG1237 Metal-dependent hydrolases of the beta-lactamase superfamily II
IYC_06953	1,689	2,354	666	Fe/Mn family superoxide dismutase	COG0605 Superoxide dismutase
IYC_06958	2,467	2,889	423	hypothetical protein	
IYC_06963	3,490	4,185	696	hypothetical protein	
IYC_06968	4,396	4,809	414	hemerythrin-like metal-binding domain-containingprotein	COG2703 Hemerythrin
IYC_06973	5,016	5,378	363	hypothetical protein	
IYC_06978	5,653	7,659	2,007	fructose-1,6-bisphosphatase	COG3855 Uncharacterized protein conserved in bacteria
IYC_06983	7,995	8,948	954	N(5)-(carboxyethyl)ornithine synthase	COG0686 Alanine dehydrogenase
IYC_06988	9,207	9,827	621	GntR family transcriptional regulator	COG2188 Transcriptional regulators
IYC_06993	10,170	10,361	192	hypothetical protein	
IYC_06998	10,651	12,006	1,356	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_07003	12,313	14,667	2,355	putative ATP-dependent helicase	COG1199 Rad3-related DNA helicases
IYC_07008	15,336	16,613	1,278	seryl-tRNA synthetase	COG0172 Seryl-tRNA synthetase
IYC_07013	17,009	17,803	795	response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_07018	18,018	20,048	2,031	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_07023	20,084	20,668	585	HDIG domain-containing protein	COG2206 HD-GYP domain
IYC_07028	20,964	22,211	1,248	response regulator	COG0784 FOG: CheY-like receiver
IYC_07033	22,498	23,442	945	alpha/beta fold family hydrolase	COG2267 Lysophospholipase
IYC_07048	26,201	27,151	951	response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_07053	27,213	27,956	744	3-ketoacyl-(acyl-carrier-protein) reductase	COG1028 Dehydrogenases with different specificities(related to short-chain alcohol dehydrogenases)
IYC_07058	28,325	28,819	495	lipid hydroperoxide peroxidase	COG2077 Peroxiredoxin
IYC_07063	28,915	29,178	264	hypothetical protein	COG0191 Fructose/tagatose bisphosphate aldolase
IYC_07068	29,397	30,050	654	tRNA (guanine-N(7)-)-methyltransferase	COG0220 Predicted S-adenosylmethionine-dependent methyltransferase
IYC_07073	30,312	31,580	1,269	hypothetical protein	
IYC_07078	31,707	32,966	1,260	arsenical pump family protein	COG1055 Na+/H+ antiporter NhaD and related arsenitepermeases
IYC_07083	33,066	33,551	486	hypothetical protein	
IYC_07088	33,662	34,810	1,149	ABC-type multidrug transport system, permease	COG0842 ABC-type multidrug transport system, permease component
IYC_07093	34,810	35,928	1,119	multidrug ABC transporter permease	COG0842 ABC-type multidrug transport system, permease component
IYC_07098	35,940	36,875	936	ABC transporter ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_07103	37,450	38,076	627	LuxR family DNA-binding response regulator	COG2197 Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain
IYC_07108	38,248	39,339	1,092	sensor histidine kinase	COG4585 Signal transduction histidine kinase
IYC_07113	39,590	40,240	651	putative metalloprotease	COG1994 Zn-dependent proteases
IYC_07118	40,483	41,625	1,143	drug ABC transporter, ATP-binding/permease protein	COG0842 ABC-type multidrug transport system, permease component

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_07123	41,627	42,751	1,125	ABC transporter permease/ATP-binding protein	COG0842 ABC-type multidrug transport system, permease component
IYC_07128	42,763	43,695	933	ABC transporter, ATP-binding protein SagG	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_07133	43,857	44,585	729	hypothetical protein	
IYC_07138	44,593	45,231	639	CAAX amino terminal protease family protein	COG1266 Predicted metal-dependent membrane protease
IYC_07143	45,206	46,564	1,359	streptolysin associated protein SagD	
IYC_07148	46,580	47,662	1,083	streptolysin associated protein SagC	COG0476 Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2
IYC_07153	47,665	48,624	960	streptolysin associated protein SagB	COG0778 Nitroreductase
IYC_07158	48,792	48,965	174	hypothetical protein	
IYC_07163	49,286	50,302	1,017	hypothetical protein	
IYC_07168	50,312	51,316	1,005	hypothetical protein	COG4632 Exopolysaccharide biosynthesis protein related to N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase
IYC_07173	51,716	53,092	1,377	PhoH family protein	COG1875 Predicted ATPase related to phosphate starvation-inducible protein PhoH
IYC_07178	53,657	53,836	180	hypothetical protein	
IYC_07183	54,377	>54557	>181	SagA protein	COG3409 Putative peptidoglycan-binding domain-containing protein; overlaps another CDS with the same product name
IYC_07188	54,546	55,607	1,062	SagA protein	COG0791 Cell wall-associated hydrolases (invasion-associated proteins); overlaps another CDS with the same product name
IYC_07193	55,925	56,107	183	hypothetical protein	
IYC_07198	56,763	58,991	2,229	cadmium-translocating P-type ATPase	COG2217 Cation transport ATPase
IYC_07203	59,067	59,432	366	putative cadmium resistance transcriptional regulator CadC	COG0640 Predicted transcriptional regulators
IYC_07208	59,647	59,808	162	hypothetical protein	
IYC_07213	59,901	60,752	852	hypothetical protein	COG1284 Uncharacterized conserved protein
IYC_07218	61,607	62,155	549	NUDIX family hydrolase	COG0494 NTP pyrophosphohydrolases including oxidative damage repair enzymes
IYC_07223	62,183	65,713	3,531	exonuclease SbcCD subunit C	COG0419 ATPase involved in DNA repair
IYC_07228	65,688	66,917	1,230	exonuclease SbcCD subunit D	COG0420 DNA repair exonuclease
Genomic Scaffold: NZ_JH470509					
IYC_07235	1	>389	>389	putative protein gp15	
Genomic Scaffold: NZ_JH470510					
IYC_07240	213	956	744	NAD-dependent deacetylase	COG0846 NAD-dependent protein deacetylases, SIR2 family
IYC_07245	1,142	2,050	909	LD-carboxypeptidase family protein	COG1619 Uncharacterized proteins, homologs of microcin C7 resistance protein MccF
IYC_07250	2,087	3,028	942	transporter, CorA metal ion transporter family protein	COG0598 Mg2+ and Co2+ transporters
IYC_07255	3,382	3,885	504	hypothetical protein	
IYC_07260	3,890	5,050	1,161	amidohydrolase-like protein	COG1473 Metal-dependent amidase/aminoacylase/carboxypeptidase
IYC_07265	5,579	5,791	213	putative CsfB protein	
IYC_07270	5,770	7,221	1,452	Orn/Lys/Arg decarboxylase	COG1982 Arginine/lysine/ornithine decarboxylases
IYC_07275	7,245	7,940	696	thymidylate kinase	COG0125 Thymidylate kinase
IYC_07280	8,021	8,350	330	hypothetical protein	COG3870 Uncharacterized protein conserved in bacteria
IYC_07285	8,366	9,310	945	DNA polymerase III subunit delta'	COG2812 DNA polymerase III, gamma/tau subunits
IYC_07290	9,312	10,226	915	PSP1 domain-containing protein	COG1774 Uncharacterized homolog of PSP1
IYC_07295	10,238	10,441	204	hypothetical protein	COG2608 Copper chaperone
IYC_07300	10,523	10,693	171	ferredoxin	COG1145 Ferredoxin
IYC_07305	10,813	11,553	741	hypothetical protein	COG4123 Predicted O-methyltransferase
IYC_07310	11,573	12,427	855	tetrapyrrole methylase family protein	COG0313 Predicted methyltransferases

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_07315	12,649	13,716	1,068	NlpC/P60 family protein	COG3883 Uncharacterized protein conserved in bacteria
IYC_07320	13,813	14,058	246	AbrB family transcriptional regulator	COG2002 Regulators of stationary/sporulation gene expression
IYC_07325	14,865	15,506	642	hypothetical protein	
IYC_07330	15,927	17,138	1,212	arginine deiminase	COG2235 Arginine deiminase
IYC_07335	17,209	17,676	468	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	COG0245 2C-methyl-D-erythritol 2,4-cyclodiphosphatesynthase
IYC_07340	18,412	19,281	870	polysaccharide deacetylase family protein	COG0726 Predicted xylanase/chitin deacetylase
IYC_07345	19,410	19,652	243	hypothetical protein	
IYC_07350	19,772	20,347	576	spore maturation protein A	COG2715 Uncharacterized membrane protein, required for spore maturation in B.subtilis.
IYC_07355	20,360	20,878	519	spore maturation protein B	COG0700 Uncharacterized membrane protein
IYC_07360	20,916	22,253	1,338	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_07365	22,397	22,969	573	uracil-DNA glycosylase	COG1573 Uracil-DNA glycosylase
IYC_07370	22,987	23,421	435	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_07375	23,801	25,738	1,938	methionyl-tRNA synthetase	COG0143 Methionyl-tRNA synthetase
IYC_07380	26,007	27,512	1,506	hypothetical protein	COG4399 Uncharacterized protein conserved in bacteria
IYC_07385	27,663	28,457	795	TatD family hydrolase	COG0084 Mg-dependent DNase
IYC_07390	28,685	29,743	1,059	3D/G5 domain-containing protein	COG3583 Uncharacterized protein conserved in bacteria
IYC_07395	29,816	30,361	546	putative primase	COG1658 Small primase-like proteins (Toprim domain)
IYC_07400	30,363	31,190	828	16S ribosomal RNA methyltransferase KsgA/Dim1 family protein	COG0030 Dimethyladenosine transferase (rRNA methylation)
IYC_07405	31,269	32,423	1,155	hypothetical protein	
IYC_07410	32,585	33,172	588	hypothetical protein	
IYC_07415	33,340	34,173	834	putative lipoprotein	
IYC_07420	34,342	35,400	1,059	RND family marcrolide efflux protein	COG0845 Membrane-fusion protein
IYC_07425	35,411	36,106	696	putative macrolide efflux ABC transporter, ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_07430	36,103	37,311	1,209	putative macrolide efflux ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_07435	37,394	38,236	843	MRP protein-like protein	COG0489 ATPases involved in chromosome partitioning
IYC_07440	38,373	38,564	192	iron-sulfur cluster-binding protein	COG1141 Ferredoxin
IYC_07445	38,995	39,171	177	hypothetical protein	
IYC_07450	39,249	40,319	1,071	putative diguanylate cyclase	COG2199 FOG: GGDEF domain
IYC_07455	40,376	41,104	729	mgtC family protein	COG1285 Uncharacterized membrane protein
IYC_07460	41,393	43,534	2,142	anaerobic ribonucleoside triphosphate reductase	COG1328 Oxygen-sensitive ribonucleoside-triphosphate reductase
IYC_07465	43,551	44,066	516	anaerobic ribonucleoside-triphosphate reductase activating protein	COG0602 Organic radical activating enzymes
IYC_07470	44,183	44,860	678	hypothetical protein	
IYC_07475	44,862	45,200	339	PadR family transcriptional regulator	COG1695 Predicted transcriptional regulators
IYC_07480	45,548	46,183	636	putative zinc-dependent hydrolase	COG2220 Predicted Zn-dependent hydrolases of the beta-lactamase fold
IYC_07485	46,625	49,741	3,117	isoleucyl-tRNA synthetase	COG0060 Isoleucyl-tRNA synthetase
IYC_07490	49,892	50,890	999	LacI family transcription regulator	COG1609 Transcriptional regulators
IYC_07495	51,107	55,165	4,059	hypothetical protein	COG1112 Superfamily I DNA and RNA helicases and helicase subunits
IYC_07500	55,242	56,309	1,068	putative mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase	COG0836 Mannose-1-phosphate guanylyltransferase
IYC_07505	56,646	57,929	1,284	hypothetical protein	COG1641 Uncharacterized conserved protein
IYC_07510	57,934	58,740	807	hypothetical protein	COG1606 ATP-utilizing enzymes of the PP-loop superfamily

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_07515	58,760	59,506	747	hypothetical protein	COG1691 NCAIR mutase (PurE)-related proteins
IYC_07520	59,731	61,113	1,383	radical SAM domain-containing protein	COG0535 Predicted Fe-S oxidoreductases
IYC_07525	61,226	61,909	684	transposase IS3/IS911 family protein	COG2963 Transposase and inactivated derivatives
IYC_07530	61,936	62,799	864	Integrase catalytic region	COG2801 Transposase and inactivated derivatives
IYC_07535	63,029	64,303	1,275	putative sugar-binding protein	COG1653 ABC-type sugar transport system, periplasmic component
IYC_07540	64,420	65,535	1,116	putative mannosyltransferase	COG0438 Glycosyltransferase
IYC_07545	65,701	66,711	1,011	CotS family spore coat protein	COG0510 Predicted choline kinase involved in LPS biosynthesis
IYC_07550	66,726	66,920	195	hypothetical protein	
IYC_07555	66,947	67,714	768	hypothetical protein	
IYC_07560	67,727	68,740	1,014	putative spore coat protein	COG0510 Predicted choline kinase involved in LPS biosynthesis
IYC_07565	68,943	70,070	1,128	glycosyl transferase, group 1 family protein	COG0438 Glycosyltransferase
IYC_07570	70,187	71,191	1,005	spore coat protein CotS	
IYC_07575	71,315	72,217	903	peptidase family protein	
IYC_07580	72,413	72,649	237	hypothetical protein	COG4466 Uncharacterized protein conserved in bacteria
IYC_07585	72,751	74,313	1,563	LysM domain-containing protein	COG1388 FOG: LysM repeat
IYC_07590	74,511	75,311	801	cyanophycinase	COG4242 Cyanophycinase and related exopeptidases
IYC_07595	75,355	77,976	2,622	cyanophycin synthetase	COG1181 D-alanine-D-alanine ligase and related ATP-grasp enzymes
IYC_07600	78,072	78,914	843	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	COG1947 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase
IYC_07605	79,074	79,271	198	hypothetical protein	
IYC_07610	79,331	80,791	1,461	spore germination protein	
IYC_07615	80,811	81,914	1,104	spore germination protein	COG2216 High-affinity K+ transport system, ATPase chain B
IYC_07620	81,958	83,136	1,179	germination protein, GerC family	
IYC_07625	83,210	83,842	633	stage II sporulation protein R	
IYC_07630	83,878	85,119	1,242	hypothetical protein	COG0707 UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase
IYC_07635	85,492	86,862	1,371	germination protein YpeB	
IYC_07640	86,989	87,969	981	hypothetical protein	COG0482 Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase, contains the PP-loop ATPase domain
IYC_07645	87,988	88,416	429	hypothetical protein	COG4506 Uncharacterized protein conserved in bacteria
IYC_07650	88,558	88,857	300	hypothetical protein	
IYC_07655	89,094	90,695	1,602	CTP synthetase	COG0504 CTP synthase (UTP-ammonia lyase)
IYC_07660	90,949	91,293	345	hypothetical protein	
IYC_07665	91,465	92,910	1,446	transcription termination factor Rho	COG1158 Transcription termination factor
IYC_07670	92,950	93,168	219	ribosomal protein L31	COG0254 Ribosomal protein L31
IYC_07675	93,394	93,969	576	thymidine kinase	COG1435 Thymidine kinase
IYC_07680	93,986	94,903	918	hypothetical protein	COG3872 Predicted metal-dependent enzyme
IYC_07685	94,911	95,762	852	N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase	COG2890 Methylase of polypeptide chain release factors
IYC_07690	95,844	96,920	1,077	peptide chain release factor 1	COG0216 Protein chain release factor A
IYC_07695	96,948	97,556	609	hypothetical protein	
IYC_07700	97,617	98,336	720	ZIP transporter family protein	COG0428 Predicted divalent heavy-metal cations transporter
IYC_07705	98,357	99,412	1,056	Sua5/YciO/YrdC/YwlC family protein	COG0009 Putative translation factor (SUA5)
IYC_07710	99,422	99,874	453	low molecular weight protein tyrosine phosphatase	COG0394 Protein-tyrosine-phosphatase
IYC_07715	99,944	100,381	438	ribose-5-phosphate isomerase B	COG0698 Ribose 5-phosphate isomerase RpiB

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_07720	100,411	101,040	630	uracil phosphoribosyltransferase	COG0035 Uracil phosphoribosyltransferase
IYC_07725	101,681	101,875	195	hypothetical protein	COG1476 Predicted transcriptional regulators
IYC_07730	101,880	102,293	414	hypothetical protein	
IYC_07735	102,360	102,536	177	hypothetical protein	
IYC_07740	102,556	102,720	165	hypothetical protein	
IYC_07745	102,830	103,084	255	hypothetical protein	COG2274 ABC-type bacteriocin/lantibiotic exporters,contain an N-terminal double-glycine peptidase domain
IYC_07750	103,294	103,662	369	hypothetical protein	
IYC_07755	103,722	103,820	99	hypothetical protein	
IYC_07760	103,848	103,979	132	hypothetical protein	
IYC_07765	104,203	106,425	2,223	ABC transporter CbaT	COG2274 ABC-type bacteriocin/lantibiotic exporters,contain an N-terminal double-glycine peptidase domain
IYC_07770	106,425	107,855	1,431	putative bacteriocin ABC transporter, bacteriocin-binding protein	COG0845 Membrane-fusion protein
IYC_07775	108,277	108,765	489	cytidine/deoxycytidylate deaminase family protein	COG2131 Deoxycytidylate deaminase
IYC_07780	108,872	109,906	1,035	undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase	COG0472 UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase
IYC_07785	109,920	111,077	1,158	UDP-N-acetylglucosamine 2-epimerase	COG0381 UDP-N-acetylglucosamine 2-epimerase
IYC_07790	111,564	111,932	369	ATP synthase F0, I subunit	
IYC_07795	111,947	112,627	681	F0F1 ATP synthase subunit A	COG0356 F0F1-type ATP synthase, subunit a
IYC_07800	112,667	112,906	240	F0F1 ATP synthase subunit C	COG0636 F0F1-type ATP synthase, subunit c/Archaeal/vacuolar-type H+-ATPase, subunit K
IYC_07805	112,986	113,465	480	F0F1 ATP synthase subunit B	COG0711 F0F1-type ATP synthase, subunit b
IYC_07810	113,468	114,007	540	F0F1 ATP synthase subunit delta	COG0712 F0F1-type ATP synthase, delta subunit (mitochondrial oligomycin sensitivity protein)
IYC_07815	114,018	115,532	1,515	F0F1 ATP synthase subunit alpha	COG0056 F0F1-type ATP synthase, alpha subunit
IYC_07820	115,554	116,402	849	F0F1 ATP synthase subunit gamma	COG0224 F0F1-type ATP synthase, gamma subunit
IYC_07825	116,416	117,804	1,389	F0F1 ATP synthase subunit beta	COG0055 F0F1-type ATP synthase, beta subunit
IYC_07830	117,835	118,236	402	F0F1 ATP synthase subunit epsilon	COG0355 F0F1-type ATP synthase, epsilon subunit (mitochondrial delta subunit)
IYC_07835	118,427	119,134	708	hypothetical protein	
IYC_07840	119,158	120,414	1,257	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	COG0766 UDP-N-acetylglucosamine enolpyruvyl transferase
IYC_07845	120,701	121,777	1,077	stage II sporulation protein D	COG2385 Sporulation protein and related proteins
IYC_07850	122,177	122,923	747	putative peptidase	COG0739 Membrane proteins related to metalloendopeptidases
IYC_07855	123,031	123,285	255	stage III sporulation protein D	
IYC_07860	123,385	124,413	1,029	rod shape-determining protein Mbl	COG1077 Actin-like ATPase involved in cell morphogenesis
IYC_07865	124,530	125,054	525	hypothetical protein	
IYC_07870	125,103	125,825	723	hypothetical protein	COG0037 Predicted ATPase of the PP-loop superfamilyimplicated in cell cycle control
IYC_07880	126,528	127,703	1,176	S-adenosylmethionine synthetase	COG0192 S-adenosylmethionine synthetase
IYC_07885	127,885	130,119	2,235	RecD/TraA family helicase	COG0507 ATP-dependent exoDNAse (exonuclease V), alpha subunit - helicase superfamily I member
IYC_07890	130,147	131,160	1,014	hypothetical protein	COG0556 Helicase subunit of the DNA excision repaircomplex
IYC_07895	131,133	131,810	678	comF protein,-like protein	COG1040 Predicted amidophosphoribosyltransferases
IYC_07900	132,236	132,763	528	ribosomal subunit interface protein	COG1544 Ribosome-associated protein Y (PSrp-1)
IYC_07905	132,996	135,503	2,508	preprotein translocase subunit SecA	COG0653 Preprotein translocase subunit SecA (ATPase, RNA helicase)
IYC_07910	135,618	135,740	123	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_07915	135,934	136,917	984	peptide chain release factor 2	COG1186 Protein chain release factor B
IYC_07920	137,212	138,564	1,353	peptidase family protein	COG2234 Predicted aminopeptidases
IYC_07925	138,672	140,828	2,157	RNA-binding protein	COG2183 Transcriptional accessory protein
IYC_07930	141,105	141,698	594	riboflavin transporter	COG3601 Predicted membrane protein
IYC_07935	142,059	143,708	1,650	hypothetical protein	
IYC_07940	143,729	143,989	261	hypothetical protein	COG2257 Uncharacterized homolog of the cytoplasmic domain of flagellar protein FhlB
IYC_07945	144,003	144,689	687	Ser/Thr protein phosphatase family protein	COG1768 Predicted phosphohydrolase
IYC_07950	144,773	145,054	282	ArsR family transcriptional regulator	COG0640 Predicted transcriptional regulators
IYC_07955	145,094	147,547	2,454	pyridine nucleotide-disulfide oxidoreductase family protein	COG0446 Uncharacterized NAD(FAD)-dependent dehydrogenases
IYC_07960	147,575	148,429	855	hypothetical protein	COG1655 Uncharacterized protein conserved in bacteria
IYC_07965	148,580	149,368	789	RelA/SpoT domain-containing protein	COG2357 Uncharacterized protein conserved in bacteria
IYC_07970	149,762	152,203	2,442	leucyl-tRNA synthetase	COG0495 Leucyl-tRNA synthetase
IYC_07975	152,626	153,030	405	hypothetical protein	COG0607 Rhodanese-related sulfurtransferase
IYC_07980	153,646	153,795	150	hypothetical protein	
IYC_07985	153,894	154,493	600	DnaJ family molecular chaperone	COG2214 DnaJ-class molecular chaperone
IYC_07990	154,480	155,355	876	hypothetical protein	
IYC_07995	155,359	155,898	540	putative adenylyl cyclase CyaB	COG1437 Adenylate cyclase, class 2 (thermophilic)
IYC_08000	156,204	157,241	1,038	hypothetical protein	
IYC_08005	157,389	158,315	927	fructose-1,6-bisphosphate aldolase, class II	COG0191 Fructose/tagatose bisphosphate aldolase
IYC_08010	158,541	158,825	285	hypothetical protein	
IYC_08015	158,942	159,610	669	hypothetical protein	COG2323 Predicted membrane protein
IYC_08020	159,733	161,391	1,659	putative threonyl-tRNA synthetase/uridine kinase	COG0441 Threonyl-tRNA synthetase
IYC_08025	161,522	161,683	162	hypothetical protein	
IYC_08030	161,753	162,274	522	CDP-diacylglycerol--serine O-phosphatidyltransferase	COG1183 Phosphatidylserine synthase
IYC_08035	162,461	162,877	417	CBS domain-containing protein	COG0517 FOG: CBS domain
IYC_08040	162,983	163,867	885	putative lipid kinase	COG1597 Sphingosine kinase and enzymes related to eukaryotic diacylglycerol kinase
IYC_08045	163,941	164,585	645	hypothetical protein	
IYC_08050	164,677	166,617	1,941	hypothetical protein	COG1032 Fe-S oxidoreductase
IYC_08055	167,215	168,480	1,266	PRC-barrel domain/MgtE domain/CBS domain-containing protein	COG2239 Mg/Co/Ni transporter MgtE (contains CBS domain)
IYC_08060	168,710	169,408	699	spore-cortex-lytic enzyme	COG3409 Putative peptidoglycan-binding domain-containing protein
IYC_08065	169,772	170,686	915	cysteine synthase A	COG0031 Cysteine synthase
IYC_08070	170,702	171,298	597	serine O-acetyltransferase	COG1045 Serine acetyltransferase
IYC_08075	171,320	172,285	966	iron-sulfur cluster-binding protein	COG1600 Uncharacterized Fe-S protein
IYC_08080	172,433	173,134	702	1-acyl-sn-glycerol-3-phosphate acyltransferase family protein	COG0204 1-acyl-sn-glycerol-3-phosphate acyltransferase
IYC_08085	173,227	173,868	642	endonuclease III	COG0177 Predicted EndoIII-related endonuclease
IYC_08090	173,907	>174010	>104	hypothetical protein	
IYC_08095	174,197	175,369	1,173	malic enzyme	COG0281 Malic enzyme
IYC_08100	175,396	176,625	1,230	competence damage-inducible protein A	COG1058 Predicted nucleotide-utilizing enzyme related to molybdopterin-biosynthesis enzyme MoeA
IYC_08105	176,651	177,634	984	hypothetical protein	COG1073 Hydrolases of the alpha/beta superfamily
IYC_08110	177,747	178,781	1,035	glutamyl aminopeptidase family protein	COG1363 Cellulase M and related proteins
IYC_08115	178,933	180,069	1,137	vanW-like family protein	COG2720 Uncharacterized vancomycin resistance protein
IYC_08120	180,270	181,622	1,353	exported protein	COG2720 Uncharacterized vancomycin resistance protein

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_08125	181,777	182,784	1,008	spore photoproduct lyase	COG1533 DNA repair photolyase
IYC_08130	183,095	183,562	468	hypothetical protein	COG0219 Predicted rRNA methylase (SpoU class)
IYC_08135	183,575	184,414	840	DegV family protein	COG1307 Uncharacterized protein conserved in bacteria
IYC_08140	184,706	185,809	1,104	bmp family lipoprotein	COG1744 Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein
IYC_08145	185,995	187,527	1,533	putative sugar ABC transporter, ATP-binding protein	COG3845 ABC-type uncharacterized transport systems,ATPase components
IYC_08150	187,529	188,632	1,104	putative sugar ABC transporter, permease protein	COG4603 ABC-type uncharacterized transport system, permease component
IYC_08155	188,625	189,551	927	putative sugar ABC transporter, permease protein	COG1079 Uncharacterized ABC-type transport system, permease component
IYC_08160	189,652	191,025	1,374	RNA polymerase factor sigma-54	COG1508 DNA-directed RNA polymerase specialized sigma subunit, sigma54 homolog
IYC_08165	191,462	192,493	1,032	central glycolytic genes regulator	COG2390 Transcriptional regulator, contains sigma factor-related N-terminal domain
IYC_08170	192,569	193,576	1,008	glyceraldehyde-3-phosphate dehydrogenase, type I	COG0057 Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase
IYC_08175	193,872	195,068	1,197	phosphoglycerate kinase	COG0126 3-phosphoglycerate kinase
IYC_08180	195,124	195,870	747	triosephosphate isomerase	COG0149 Triosephosphate isomerase
IYC_08185	195,954	197,483	1,530	phosphoglyceromutase	COG0696 Phosphoglyceromutase
IYC_08190	197,539	198,834	1,296	enolase	COG0148 Enolase
IYC_08195	199,102	199,335	234	preprotein translocase, SecG subunit	COG1314 Preprotein translocase subunit SecG
IYC_08200	199,544	201,703	2,160	ribonuclease R	COG0557 Exoribonuclease R
IYC_08205	202,102	202,860	759	stationary phase survival protein SurE	COG0496 Predicted acid phosphatase
IYC_08210	203,052	203,522	471	SsrA-binding protein	COG0691 tmRNA-binding protein
IYC_08215	203,678	204,841	1,164	sodium:dicarboxylate symporter family protein	COG1301 Na+/H+-dicarboxylate symporters
IYC_08220	206,278	206,688	411	hypothetical protein	
IYC_08225	207,091	207,822	732	transcriptional regulator, AraC family protein	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_08230	207,892	208,047	156	hypothetical protein	
IYC_08235	208,070	208,174	105	hypothetical protein	
IYC_08250	209,115	209,966	852	flagellin protein	COG1344 Flagellin and related hook-associated proteins
IYC_08255	210,436	211,077	642	hypothetical protein	COG0662 Mannose-6-phosphate isomerase
IYC_08260	211,319	212,002	684	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_08265	212,071	213,075	1,005	periplasmic sensor signal transduction histidinekinase	COG0642 Signal transduction histidine kinase
IYC_08270	213,150	213,914	765	bacitracin export ATP-binding protein BceA	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_08275	213,901	215,823	1,923	hypothetical protein	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_08280	215,816	216,787	972	hypothetical protein	COG0596 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
IYC_08285	217,074	217,286	213	hypothetical protein	
IYC_08290	217,391	217,891	501	hypothetical protein	
IYC_08295	218,225	218,752	528	RNA polymerase ecf-type sigma factor	COG1595 DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
IYC_08300	218,754	220,151	1,398	hypothetical protein	
IYC_08305	220,560	221,669	1,110	hypothetical protein	
IYC_08310	222,262	222,759	498	acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_08315	223,068	223,592	525	riboflavin biosynthesis protein RibD domain-containing protein	COG0262 Dihydrofolate reductase
IYC_08320	223,655	224,578	924	DNA polymerase III subunit epsilon	COG0847 DNA polymerase III, epsilon subunit and related 3'-5' exonucleases

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_08325	225,155	225,625	471	hypothetical protein	COG1879 ABC-type sugar transport system, periplasmic component
IYC_08330	225,636	226,319	684	putative cytochrome c-type biogenesis protein CcdA	COG0785 Cytochrome c biogenesis protein
IYC_08335	226,341	226,907	567	putative thiol-disulfide oxidoreductase ResA	COG0526 Thiol-disulfide isomerase and thioredoxins
IYC_08340	226,932	227,609	678	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_08345	227,602	228,999	1,398	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_08350	229,201	230,523	1,323	putative oxidoreductase, FAD-binding protein	COG0277 FAD/FMN-containing dehydrogenases
IYC_08355	230,598	>230706	>109	hypothetical protein	
Genomic Scaffold: NZ_JH470512					
IYC_08368	<1	152	>152	hypothetical protein	
IYC_08373	225	>370	>146	transposase	COG2801 Transposase and inactivated derivatives
Genomic Scaffold: NZ_JH470513					
IYC_08378	<1	>213	>213	integrase	COG2801 Transposase and inactivated derivatives
Genomic Scaffold: NZ_JH470514					
IYC_08383	<1	134	>134	hypothetical protein	
IYC_08388	167	910	744	peptidil-prolyl isomerase family protein	COG0760 Parvulin-like peptidyl-prolyl isomerase
Genomic Scaffold: NZ_JH470515					
IYC_08393	<1	96	>96	hypothetical protein	
IYC_08398	142	>259	>118	hypothetical protein	
Genomic Scaffold: NZ_JH470516					
IYC_08403	<1	233	>233	putative protein gp15	
IYC_08408	233	577	345	hypothetical protein	
IYC_08413	577	1,086	510	hypothetical protein	
IYC_08418	1,087	1,521	435	hypothetical protein	
IYC_08423	1,518	1,862	345	hypothetical protein	
IYC_08428	1,862	2,272	411	hypothetical protein	
IYC_08433	2,274	2,534	261	hypothetical protein	
IYC_08438	2,544	2,714	171	hypothetical protein	
IYC_08443	2,760	3,659	900	major head protein	
IYC_08448	3,677	4,270	594	hypothetical protein	
IYC_08453	4,330	5,049	720	hypothetical protein	COG5585 NAD+--asparagine ADP-ribosyltransferase
IYC_08458	5,049	6,320	1,272	SPP1 family phage portal protein	
IYC_08463	6,324	8,099	1,776	hypothetical protein	
IYC_08468	8,164	8,826	663	resolvase	COG1961 Site-specific recombinases, DNA invertase Pin homologs
IYC_08473	9,075	9,488	414	DNA-binding protein	COG3415 Transposase and inactivated derivatives
IYC_08478	9,572	10,084	513	hypothetical protein	COG3013 Uncharacterized conserved protein
IYC_08483	10,283	10,993	711	hypothetical protein	
IYC_08488	11,580	11,732	153	hypothetical protein	
IYC_08493	11,873	12,358	486	hypothetical protein	
IYC_08498	12,404	12,571	168	hypothetical protein	
IYC_08503	12,574	12,975	402	hypothetical protein	
IYC_08508	13,070	13,285	216	hypothetical protein	
IYC_08513	13,317	13,493	177	hypothetical protein	
IYC_08518	13,575	13,733	159	hypothetical protein	COG1396 Predicted transcriptional regulators
IYC_08523	13,735	13,872	138	hypothetical protein	
IYC_08528	13,884	14,210	327	hypothetical protein	
IYC_08533	14,210	14,611	402	endodeoxyribonuclease RusA family protein	COG4570 Holliday junction resolvase
IYC_08538	14,608	14,772	165	hypothetical protein	
IYC_08543	14,801	15,001	201	hypothetical protein	
IYC_08548	15,054	15,236	183	hypothetical protein	
IYC_08553	15,274	15,552	279	nucleotide pyrophosphohydrolase domain-containing protein	
IYC_08558	15,530	16,057	528	hypothetical protein	
IYC_08563	16,058	16,318	261	hypothetical protein	
IYC_08568	16,321	16,863	543	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_08573	16,864	17,652	789	phage replication initiation protein	
IYC_08578	17,652	17,939	288	hypothetical protein	
IYC_08583	17,950	18,570	621	essential recombination function protein	
IYC_08588	18,570	19,067	498	phage protein	
IYC_08593	19,040	19,198	159	hypothetical protein	
IYC_08598	19,201	>19440	>240	hypothetical protein	
Genomic Scaffold: NZ_JH470517					
IYC_08603	<1	125	>125	hypothetical protein	
IYC_08608	253	>626	>374	hypothetical protein	
Genomic Scaffold: NZ_JH470518					
IYC_08623	1,348	1,524	177	hypothetical protein	
IYC_08628	1,546	>1671	>126	hypothetical protein	
Genomic Scaffold: NZ_JH470519					
IYC_08633	144	281	138	hypothetical protein	
IYC_08638	1,233	2,051	819	hypothetical protein	
IYC_08643	2,072	2,698	627	hypothetical protein	
IYC_08648	3,020	4,276	1,257	pyridine nucleotide-disulfide oxidoreductase	COG0446 Uncharacterized NAD(FAD)-dependent dehydrogenases
IYC_08653	4,654	5,277	624	endodeoxyribonuclease RusA family protein	COG4570 Holliday junction resolvase
IYC_08658	5,501	6,523	1,023	UGMP family protein	COG0533 Metal-dependent proteases with possible chaperone activity
IYC_08663	6,983	10,417	3,435	pyruvate carboxylase	COG1038 Pyruvate carboxylase
IYC_08668	10,797	13,724	2,928	peptidase family protein	COG1026 Predicted Zn-dependent peptidases, insulinase-like
IYC_08673	13,870	15,288	1,419	carbon starvation protein CstA	COG1966 Carbon starvation protein, predicted membrane protein
IYC_08678	15,531	17,165	1,635	sensor histidine kinase	COG3275 Putative regulator of cell autolysis
IYC_08683	17,166	17,930	765	LytTr family DNA-binding response regulator	COG3279 Response regulator of the LytR/AlgR family
IYC_08688	18,121	20,043	1,923	putative drug resistance ABC transporter, ATP-binding protein	COG0488 ATPase components of ABC transporters with duplicated ATPase domains
IYC_08693	20,243	20,875	633	redox-sensing transcriptional repressor Rex	COG2344 AT-rich DNA-binding protein
IYC_08698	21,105	21,995	891	ferredoxin-NADP(+) reductase subunit alpha	COG0543 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases
IYC_08703	21,995	23,380	1,386	dihydropyrimidine dehydrogenase subunit A	COG0493 NADPH-dependent glutamate synthase beta chain and related oxidoreductases
IYC_08708	23,530	24,447	918	putative 8-oxoguanine DNA glycosylase	COG0122 3-methyladenine DNA glycosylase/8-oxoguanine DNA glycosylase
IYC_08713	24,621	26,114	1,494	[Fe] hydrogenase	COG1145 Ferredoxin
IYC_08718	26,362	27,096	735	hypothetical protein	COG0398 Uncharacterized conserved protein
IYC_08723	27,091	27,735	645	integral membrane protein	COG4478 Predicted membrane protein
IYC_08728	28,050	28,337	288	co-chaperonin GroES	COG0234 Co-chaperonin GroES (HSP10)
IYC_08733	28,366	29,991	1,626	chaperonin GroEL	COG0459 Chaperonin GroEL (HSP60 family)
IYC_08738	30,064	30,813	750	PP-loop family protein	COG0037 Predicted ATPase of the PP-loop superfamilyimplicated in cell cycle control
IYC_08743	31,255	32,709	1,455	inosine 5'-monophosphate dehydrogenase	COG0516 IMP dehydrogenase/GMP reductase
IYC_08748	32,722	34,254	1,533	GMP synthase	COG0518 GMP synthase - Glutamine amidotransferase domain
IYC_08753	34,709	36,031	1,323	branched-chain amino acid transport system II carrier protein	COG1114 Branched-chain amino acid permeases
IYC_08758	36,497	38,128	1,632	AMP-binding enzyme	COG0318 Acyl-CoA synthetases (AMP-forming)/AMP-acidligases II
IYC_08763	38,183	39,421	1,239	E-cinnamoyl-CoA:R-phenyllactate CoA transferase	COG1804 Predicted acyl-CoA transferases/carnitine dehydratase
IYC_08768	39,421	40,215	795	R-phenyllactate dehydratase activator	COG1924 Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)
IYC_08773	40,220	41,443	1,224	R-phenyllactate dehydratase subunit B	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_08778	41,445	42,569	1,125	R-phenyllactate dehydratase, C subunit	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB
IYC_08783	42,735	43,868	1,134	acyl-CoA dehydrogenase family protein	COG1960 Acyl-CoA dehydrogenases
IYC_08788	43,881	44,675	795	electron transfer flavoprotein subunit beta/FixAfamily protein	COG2086 Electron transfer flavoprotein, beta subunit
IYC_08793	44,693	45,889	1,197	electron transfer flavoprotein, alpha subunit/FixB family protein	COG1145 Ferredoxin
IYC_08798	46,096	47,271	1,176	hypothetical protein	COG3949 Uncharacterized membrane protein
IYC_08803	48,170	49,165	996	D-lactate dehydrogenase	COG1052 Lactate dehydrogenase and related dehydrogenases
IYC_08808	49,345	50,349	1,005	D-lactate dehydrogenase	COG1052 Lactate dehydrogenase and related dehydrogenases
IYC_08813	50,882	51,526	645	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_08818	51,850	52,020	171	hypothetical protein	
IYC_08823	52,104	53,456	1,353	glucose-6-phosphate isomerase	COG0166 Glucose-6-phosphate isomerase
IYC_08828	53,467	53,901	435	hypothetical protein	COG1671 Uncharacterized protein conserved in bacteria
IYC_08833	54,053	54,220	168	hypothetical protein	
IYC_08838	54,354	56,357	2,004	triple tyrosine motif-containing protein	
IYC_08843	56,416	57,087	672	Slt family transglycosylase	COG0741 Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)
IYC_08848	57,172	57,888	717	RNA pseudouridine synthase family protein	COG1187 16S rRNA uridine-516 pseudouridylate synthase and related pseudouridylate synthases
IYC_08853	58,185	58,484	300	YerC/YecD family protein	COG4496 Uncharacterized protein conserved in bacteria
IYC_08858	58,612	60,828	2,217	ATP-dependent DNA helicase PcrA	COG0210 Superfamily I DNA and RNA helicases
IYC_08863	60,840	62,834	1,995	NAD-dependent DNA ligase LigA	COG0272 NAD-dependent DNA ligase (contains BRCT domain type II)
IYC_08868	62,935	63,204	270	hypothetical protein	
IYC_08873	63,593	64,804	1,212	sodium:dicarboxylate symporter family protein	COG1301 Na+/H+-dicarboxylate symporters
IYC_08878	64,941	65,228	288	aspartyl/glutamyl-tRNA amidotransferase subunit C	COG0721 Asp-tRNAAsn/Glu-tRNAGln amidotransferase C subunit
IYC_08883	65,265	66,722	1,458	aspartyl/glutamyl-tRNA amidotransferase subunit A	COG0154 Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases
IYC_08888	66,768	68,198	1,431	aspartyl/glutamyl-tRNA amidotransferase subunit B	COG0064 Asp-tRNAAsn/Glu-tRNAGln amidotransferase B subunit (PET112 homolog)
IYC_08893	68,353	69,915	1,563	solute-binding family 5 protein	COG4166 ABC-type oligopeptide transport system, periplasmic component
IYC_08898	69,918	70,073	156	hypothetical protein	
IYC_08903	70,186	71,559	1,374	hypothetical protein	
IYC_08908	71,663	72,262	600	dITP/XTP pyrophosphatase	COG0127 Xanthosine triphosphate pyrophosphatase
IYC_08913	72,279	72,749	471	phosphodiesterase	COG0622 Predicted phosphoesterase
Genomic Scaffold: NZ_JH470520					
IYC_08944	<1	685	>685	putative electron transport protein	COG4656 Predicted NADH:ubiquinone oxidoreductase, subunit RnfC
IYC_08949	1,553	2,449	897	hypothetical protein	
Genomic Scaffold: NZ_JH470521					
IYC_08954	75	569	495	hypothetical protein	
IYC_08959	602	763	162	hypothetical protein	
IYC_08964	813	>1036	>224	phage protein	
Genomic Scaffold: NZ_JH470522					
IYC_08969	63	554	492	hypothetical protein	
IYC_08974	696	1,310	615	hypothetical protein	
IYC_08979	1,312	1,629	318	PadR family transcriptional regulator	COG1695 Predicted transcriptional regulators

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_08984	1,849	2,703	855	GNAT family acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_08989	2,897	3,229	333	TfoX domain-containing protein	COG3070 Regulator of competence-specific genes
IYC_08994	3,448	4,290	843	aldo/keto reductase family oxidoreductase	COG0656 Aldo/keto reductases, related to diketogulonate reductase
IYC_08999	4,296	4,661	366	hypothetical protein	
IYC_09004	4,779	5,372	594	hypothetical protein	COG3506 Uncharacterized conserved protein
IYC_09009	5,541	5,984	444	acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_09014	6,191	6,655	465	acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_09019	6,809	6,967	159	hypothetical protein	
IYC_09024	7,065	7,526	462	hypothetical protein	
IYC_09029	7,991	9,028	1,038	AraC family transcriptional regulator	COG4936 Predicted sensor domain
IYC_09034	9,649	10,779	1,131	glycerol dehydrogenase	COG0371 Glycerol dehydrogenase and related enzymes
IYC_09039	10,790	12,550	1,761	dihydroxyacetone kinase	COG2376 Dihydroxyacetone kinase
IYC_09044	13,093	14,115	1,023	putative iron chelate uptake ABC transporter, solute-binding protein	COG0614 ABC-type Fe3+-hydroxamate transport system,periplasmic component
IYC_09049	14,127	15,194	1,068	iron chelate ABC transporter permease protein	COG0609 ABC-type Fe3+-siderophore transport system,permease component
IYC_09054	15,195	15,965	771	iron chelate ABC transporter ATP-binding protein	COG1120 ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components
IYC_09059	16,503	17,924	1,422	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_09064	17,914	18,636	723	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_09069	19,150	19,926	777	hypothetical protein	
IYC_09074	19,930	20,817	888	ABC transporter ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_09079	20,838	21,857	1,020	hypothetical protein	
IYC_09084	22,166	23,125	960	Abortive infection protein	COG1266 Predicted metal-dependent membrane protease
IYC_09089	23,380	24,408	1,029	hypothetical protein	
IYC_09094	24,849	25,601	753	cell wall-associated hydrolase-like protein	COG3863 Uncharacterized distant relative of cell wall-associated hydrolases
IYC_09099	25,835	26,182	348	hypothetical protein	
IYC_09104	26,330	27,352	1,023	molybdopterin biosynthesis protein	COG0303 Molybdopterin biosynthesis enzyme
IYC_09109	28,102	29,457	1,356	two component system histidine kinase	COG0642 Signal transduction histidine kinase
IYC_09114	29,441	30,127	687	regulatory protein VanR	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_09119	30,311	32,641	2,331	efflux ABC transporter, permease protein	
IYC_09124	32,628	33,296	669	macrolide export ATP-binding/permease protein MacB	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_09129	33,446	34,411	966	sulfite reductase, subunit C	COG2221 Dissimilatory sulfite reductase (desulfoviridin), alpha and beta subunits
IYC_09134	34,424	35,221	798	anaerobic sulfite reductase subunit B	COG0543 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases
IYC_09139	35,208	36,236	1,029	anaerobic sulfite reductase subunit A	COG1145 Ferredoxin
IYC_09144	36,256	37,020	765	formate/nitrite transporter family protein	COG2116 Formate/nitrite family of transporters
IYC_09149	37,120	37,809	690	Crp/Fnr family transcriptional regulator	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_09154	37,932	38,453	522	cob(I)alamin adenolsyltransferase/cobinamide ATP-dependent adenolsyltransferase	COG2109 ATP:corrinoid adenosyltransferase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_09159	38,714	39,553	840	putative kinase	COG4542 Protein involved in propanediol utilization, and related proteins (includes coumermycin biosynthetic protein), possible kinase
IYC_09164	39,566	40,348	783	anaerobic sirohydrochlorin cobaltochelatase	COG4822 Cobalamin biosynthesis protein CbiK, Co2+ chelatase
IYC_09169	40,524	41,012	489	hypothetical protein	COG3601 Predicted membrane protein
IYC_09174	41,077	41,181	105	hypothetical protein	
IYC_09179	41,355	42,122	768	cobalt-precorrin-6x reductase	COG2099 Precorrin-6x reductase
IYC_09184	42,297	43,028	732	precorrin-3B C(17)-methyltransferase	COG1010 Precorrin-3B methylase
IYC_09189	43,308	44,747	1,440	CbiG	COG2073 Cobalamin biosynthesis protein CbiG
IYC_09194	44,784	45,548	765	precorrin-4 C11-methyltransferase	COG2875 Precorrin-4 methylase
IYC_09199	45,707	46,423	717	cobalt-precorrin-2 C(20)-methyltransferase	COG2243 Precorrin-2 methylase
IYC_09204	46,489	46,830	342	hypothetical protein	COG0222 Ribosomal protein L7/L12
IYC_09209	46,891	47,352	462	hypothetical protein	
IYC_09214	47,368	47,631	264	hypothetical protein	
IYC_09219	48,130	48,681	552	hypothetical protein	
IYC_09224	48,717	49,148	432	hypothetical protein	
IYC_09229	49,141	49,722	582	precorrin-6y C5,15-methyltransferase subunit CbiT	COG2242 Precorrin-6B methylase 2
IYC_09234	49,835	50,290	456	hypothetical protein	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_09239	50,331	50,978	648	precorrin-6y C5,15-methyltransferase (decarboxylating), CbiE subunit	COG2241 Precorrin-6B methylase 1
IYC_09244	51,205	52,284	1,080	cobalt-precorrin-6A synthase	COG1903 Cobalamin biosynthesis protein CbiD
IYC_09249	52,297	52,740	444	hypothetical protein	
IYC_09254	52,827	53,360	534	hypothetical protein	COG2320 Uncharacterized conserved protein
IYC_09259	53,467	53,763	297	membrane protein	
IYC_09264	53,833	54,354	522	nitroreductase family protein	COG0778 Nitroreductase
IYC_09269	54,435	55,496	1,062	putative membrane protein	
IYC_09274	55,787	56,215	429	GNAT family acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_09279	56,516	57,124	609	HD domain-containing protein	
IYC_09284	57,424	58,050	627	precorrin-8X methylmutase	COG2082 Precorrin isomerase
IYC_09289	58,177	58,761	585	hexapeptide repeat-containing transferase	COG0110 Acetyltransferase (isoleucine patch superfamily)
IYC_09294	58,823	59,518	696	hypothetical protein	COG0655 Multimeric flavodoxin WrbA
IYC_09299	59,617	59,931	315	hypothetical protein	
IYC_09304	60,026	61,348	1,323	cobyrinic acid a,c-diamide synthase	COG1797 Cobyrinic acid a,c-diamide synthase
IYC_09309	61,581	62,555	975	delta-aminolevulinic acid dehydratase	COG0113 Delta-aminolevulinic acid dehydratase
IYC_09314	62,646	64,106	1,461	uroporphyrinogen III synthase/methyltransferase	COG0007 Uroporphyrinogen-III methylase
IYC_09319	64,484	65,356	873	porphobilinogen deaminase	COG0181 Porphobilinogen deaminase
IYC_09324	65,325	65,972	648	precorrin-2 dehydrogenase	COG1648 Siroheme synthase (precorrin-2 oxidase/ferrochelatase domain)
IYC_09329	66,423	67,121	699	cobalt ABC transporter, permease protein CbiQ	COG0619 ABC-type cobalt transport system, permease component CbiQ and related transporters
IYC_09334	67,084	67,383	300	cobalt transport protein	COG1930 ABC-type cobalt transport system, periplasmic component
IYC_09339	67,387	68,106	720	cobalt transport protein CbiM	COG0310 ABC-type Co2+ transport system, permease component
IYC_09344	68,370	69,458	1,089	aminotransferase family protein	COG0079 Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase
IYC_09349	69,745	70,716	972	cobalamin biosynthesis protein	COG1270 Cobalamin biosynthesis protein CobD/CbiB
IYC_09354	70,718	72,199	1,482	cobyric acid synthase	COG1492 Cobyric acid synthase
IYC_09359	72,399	72,530	132	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_09364	72,794	73,468	675	catabolite gene activator and regulatory subunitof cAMP-dependent protein kinase	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_09369	73,616	74,050	435	hypothetical protein	COG2105 Uncharacterized conserved protein
IYC_09374	74,050	74,694	645	pyrrolidone-carboxylate peptidase	COG2039 Pyrrolidone-carboxylate peptidase (N-terminal pyroglutamyl peptidase)
IYC_09379	74,914	75,843	930	hypothetical protein	COG3817 Predicted membrane protein
IYC_09384	75,856	76,527	672	hypothetical protein	COG3819 Predicted membrane protein
IYC_09389	76,832	77,599	768	transcriptional activator TipA	COG0789 Predicted transcriptional regulators
IYC_09394	77,739	78,230	492	hypothetical protein	
IYC_09399	78,239	79,165	927	hypothetical protein	
IYC_09404	79,254	79,811	558	adenosylcobinamide kinase/adenosylcobinamide-phosphate guanylyltransferase	COG2087 Adenosyl cobinamide kinase/adenosyl cobinamide phosphate guanylyltransferase
IYC_09409	80,153	81,601	1,449	alpha/beta fold family hydrolase	COG1075 Predicted acetyltransferases and hydrolaseswith the alpha/beta hydrolase fold
IYC_09414	81,907	83,514	1,608	putative LexA repressor	COG0210 Superfamily I DNA and RNA helicases
IYC_09419	83,774	84,007	234	hypothetical protein	
IYC_09424	84,350	84,997	648	cyclase family protein	COG1878 Predicted metal-dependent hydrolase
IYC_09429	85,173	85,829	657	hypothetical protein	COG0247 Fe-S oxidoreductase
IYC_09434	86,221	87,933	1,713	hypothetical protein	COG4096 Type I site-specific restriction-modification system, R (restriction) subunit and related helicases
IYC_09439	88,170	88,718	549	BioY family protein	COG1268 Uncharacterized conserved protein
IYC_09444	89,030	90,055	1,026	branched-chain amino acid aminotransferase	COG0115 Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase
IYC_09449	90,447	91,475	1,029	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_09454	91,475	92,146	672	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_09459	92,306	94,666	2,361	ABC transporter, ATP-binding/permease protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_09464	94,779	96,386	1,608	hypothetical protein	
IYC_09469	96,632	97,192	561	DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_09474	97,358	98,242	885	drug/metabolite exporter family protein	COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily
IYC_09479	98,708	99,145	438	hypothetical protein	
IYC_09484	99,538	100,986	1,449	putative thiosulfate sulfurtransferase	COG2897 Rhodanese-related sulfurtransferase
IYC_09489	101,025	101,447	423	hypothetical protein	
IYC_09494	102,083	103,108	1,026	ornithine cyclodeaminase	COG2423 Predicted ornithine cyclodeaminase, mu-crystallin homolog
IYC_09499	103,290	104,021	732	hypothetical protein	
IYC_09504	104,530	104,709	180	acetyltransferase	COG3981 Predicted acetyltransferase
IYC_09509	104,839	105,243	405	hypothetical protein	
IYC_09514	105,755	106,099	345	hypothetical protein	
IYC_09519	106,423	106,800	378	ErfK/YbiS/YcfS/YnhG family protein	COG1376 Uncharacterized protein conserved in bacteria
IYC_09524	106,881	107,468	588	hypothetical protein	COG2320 Uncharacterized conserved protein
IYC_09529	107,872	108,552	681	hypothetical protein	COG3861 Uncharacterized protein conserved in bacteria
IYC_09534	108,631	109,095	465	hypothetical protein	COG3861 Uncharacterized protein conserved in bacteria
IYC_09539	109,372	110,232	861	transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_09544	110,280	111,512	1,233	peptidyl-arginine deiminase	COG2957 Peptidylarginine deiminase and related enzymes
IYC_09559	113,880	114,674	795	hypothetical protein	COG3191 L-aminopeptidase/D-esterase
IYC_09564	114,856	116,661	1,806	multicopper oxidase family protein	COG2132 Putative multicopper oxidases
IYC_09569	116,957	117,826	870	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_09574	117,899	119,011	1,113	cholovlglycine hydrolase	COG4927 Predicted cholovlglycine hydrolase
IYC_09579	119,264	119,962	699	propanediol utilization protein PduB	COG4816 Ethanolamine utilization protein
IYC_09584	120,256	120,459	204	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_09589	120,467	120,940	474	putative nimA protein	COG3467 Predicted flavin-nucleotide-binding protein
IYC_09594	121,346	121,630	285	membrane protein	
IYC_09599	122,001	122,480	480	pyridoxamine 5'-phosphate oxidase family protein	COG3467 Predicted flavin-nucleotide-binding protein
IYC_09604	122,569	122,664	96	hypothetical protein	
IYC_09609	122,939	124,348	1,410	GntR family transcriptional regulator	COG1167 Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs
IYC_09614	124,410	125,501	1,092	radical SAM domain-containing protein	COG0535 Predicted Fe-S oxidoreductases
IYC_09619	125,523	126,086	564	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_09634	127,117	127,788	672	transcriptional regulator, Crp/Fnr family protein	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_09639	127,887	128,369	483	permease-like protein	COG0701 Predicted permeases
IYC_09644	128,366	129,091	726	hypothetical protein	COG1141 Ferredoxin
IYC_09649	129,309	129,515	207	hypothetical protein	
IYC_09654	129,691	129,876	186	hypothetical protein	
IYC_09659	129,913	130,119	207	hypothetical protein	
IYC_09664	130,577	131,095	519	nitroreductase family protein	COG0778 Nitroreductase
IYC_09669	131,462	132,625	1,164	MerR-family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_09674	133,029	133,793	765	hypothetical protein	COG2966 Uncharacterized conserved protein
IYC_09679	133,793	134,263	471	hypothetical protein	COG3610 Uncharacterized conserved protein
IYC_09684	134,740	135,507	768	short-chain dehydrogenase/reductase SDR	COG0300 Short-chain dehydrogenases of various substrate specificities
IYC_09689	135,573	136,157	585	Crp/Fnr family transcriptional regulator	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_09694	136,291	136,554	264	hypothetical protein	
IYC_09699	136,716	137,120	405	putative acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_09714	138,269	139,711	1,443	lipase	COG1075 Predicted acetyltransferases and hydrolaseswith the alpha/beta hydrolase fold
IYC_09719	140,426	141,169	744	putative drug resistance ABC-2 type transporter,permease protein	COG0842 ABC-type multidrug transport system, permease component
IYC_09724	141,156	141,887	732	ABC transporter ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_09729	141,939	142,703	765	MerR family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_09734	143,053	143,577	525	acetyltransferase	COG3153 Predicted acetyltransferase
IYC_09739	143,677	143,826	150	hypothetical protein	
IYC_09744	143,823	144,098	276	spore germination protein	overlaps another CDS with the same product name
IYC_09749	144,055	145,077	1,023	spore germination protein	overlaps another CDS with the same product name
IYC_09754	145,167	145,295	129	hypothetical protein	
IYC_09759	145,510	145,755	246	hypothetical protein	COG0021 Transketolase
IYC_09764	145,786	145,968	183	hypothetical protein	
IYC_09769	146,123	146,563	441	hypothetical protein	
IYC_09774	146,710	146,904	195	hypothetical protein	COG1476 Predicted transcriptional regulators
IYC_09779	146,894	147,112	219	hypothetical protein	
IYC_09784	148,137	149,258	1,122	putative pleiotropic regulatory protein	COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis
IYC_09789	149,520	150,695	1,176	exopolysaccharide biosynthesis protein	COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis
IYC_09794	150,750	152,084	1,335	MerR family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_09799	152,528	152,989	462	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	COG0801 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_09804	153,292	154,479	1,188	dihydropteroate synthase	COG0294 Dihydropteroate synthase and related enzymes
IYC_09809	154,751	155,104	354	putative 6-pyruvoyl tetrahydropterin synthase	COG0720 6-pyruvoyl-tetrahydropterin synthase
IYC_09814	155,378	156,730	1,353	hypothetical protein	COG5298 Uncharacterized protein conserved in bacteria
IYC_09819	157,235	157,483	249	hypothetical protein	
IYC_09824	157,501	157,683	183	hypothetical protein	
IYC_09829	157,818	158,102	285	hypothetical protein	
IYC_09834	158,519	159,307	789	hypothetical protein	
IYC_09839	159,333	160,079	747	cobalamin synthase	COG0368 Cobalamin-5-phosphate synthase
IYC_09844	160,291	161,856	1,566	vanW family protein	COG2720 Uncharacterized vancomycin resistance protein
IYC_09849	162,174	162,788	615	alpha-ribazole phosphatase	COG0406 Fructose-2,6-bisphosphatase
IYC_09854	163,524	165,047	1,524	ferrous iron transport protein B	COG0370 Fe2+ transport system protein B
IYC_09859	165,049	165,825	777	ferrous iron transport protein B	COG0370 Fe2+ transport system protein B
IYC_09864	165,964	166,200	237	ferrous iron transport protein A	COG1918 Fe2+ transport system protein A
IYC_09869	166,422	168,899	2,478	methyl-accepting chemotaxis protein/ extracellular solute-binding protein family 5	COG0840 Methyl-accepting chemotaxis protein
IYC_09874	169,914	171,221	1,308	branched-chain amino acid transport system II carrier protein	COG1114 Branched-chain amino acid permeases
IYC_09879	171,475	172,692	1,218	branched-chain amino acid transport system II carrier protein	COG1114 Branched-chain amino acid permeases
IYC_09884	173,256	174,425	1,170	putative amidohydrolase	COG1228 Imidazolonepropionase and related amidohydrolases
IYC_09889	174,763	176,160	1,398	hypothetical protein	
IYC_09894	176,376	176,489	114	hypothetical protein	
IYC_09899	176,614	177,135	522	isochorismatase	COG1335 Amidases related to nicotinamidase
IYC_09904	177,232	177,726	495	putative acetyltransferase	COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins
IYC_09909	177,779	178,618	840	sulfurtransferase family protein	COG2897 Rhodanese-related sulfurtransferase
IYC_09914	178,957	179,754	798	dihydrodipicolinate reductase	COG0289 Dihydrodipicolinate reductase
IYC_09919	180,450	181,232	783	hypothetical protein	
IYC_09924	181,561	181,749	189	small acid-soluble spore protein, H-type	
IYC_09929	181,901	182,542	642	hypothetical protein	COG1139 Uncharacterized conserved protein containing a ferredoxin-like domain
IYC_09934	182,872	184,134	1,263	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_09939	184,306	185,019	714	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_09944	185,227	185,346	120	hypothetical protein	
IYC_09949	185,726	188,422	2,697	putative ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_09954	188,409	189,101	693	ABC transporter ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_09959	189,617	190,624	1,008	hypothetical protein	
IYC_09964	190,854	191,849	996	hypothetical protein	
IYC_09969	192,695	192,988	294	hypothetical protein	
IYC_09974	193,317	194,441	1,125	hypothetical protein	COG5438 Predicted multitransmembrane protein
IYC_09979	194,576	195,205	630	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_09984	195,251	196,108	858	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_09999	197,049	198,731	1,683	ABC transporter permease	
IYC_10004	198,734	199,522	789	ABC transporter, ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_10009	199,935	200,399	465	putative transporter	COG2426 Predicted membrane protein
IYC_10014	200,631	200,861	231	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_10019	201,064	201,879	816	phosphomethylpyrimidine kinase	COG0351 Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase
IYC_10024	202,008	203,222	1,215	putative thiaminase I	COG1653 ABC-type sugar transport system, periplasmic component
IYC_10029	203,268	204,026	759	UbiE/COQ5 family methyltransferase	COG2226 Methylase involved in ubiquinone/menaquinone biosynthesis
IYC_10034	204,038	205,012	975	thymidylate synthase	COG0207 Thymidylate synthase
IYC_10039	205,002	205,463	462	hypothetical protein	
IYC_10044	205,866	207,650	1,785	multidrug resistance ABC transporter ATP-bindingand permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_10049	207,650	208,252	603	transcriptional regulator	COG1846 Transcriptional regulators
IYC_10054	208,422	208,550	129	hypothetical protein	
IYC_10059	208,956	209,504	549	phosphoglycerate mutase family protein	COG0406 Fructose-2,6-bisphosphatase
IYC_10064	210,585	211,658	1,074	arsenical-resistance protein	COG0798 Arsenite efflux pump ACR3 and related permeases
IYC_10069	211,757	213,502	1,746	arsenical pump-driving ATPase ArsA	COG0003 Oxyanion-translocating ATPase
IYC_10074	213,521	213,892	372	arsenical resistance operon repressor ArsD	
IYC_10079	213,914	214,237	324	arsenical resistance operon repressor ArsR	COG0640 Predicted transcriptional regulators
IYC_10084	214,737	215,153	417	arsenate reductase family protein	COG0394 Protein-tyrosine-phosphatase
IYC_10089	215,362	215,589	228	hypothetical protein	
IYC_10094	215,610	215,831	222	hypothetical protein	
IYC_10099	216,011	216,190	180	hypothetical protein	
IYC_10104	216,347	216,646	300	hypothetical protein	
IYC_10109	216,804	217,343	540	sigma-70 family RNA polymerase sigma factor	COG1191 DNA-directed RNA polymerase specialized sigma subunit
IYC_10114	217,362	218,816	1,455	putative dnaD protein	COG3935 Putative primosome component and related proteins
IYC_10119	219,054	219,254	201	DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_10124	219,398	219,820	423	putative DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_10129	219,913	220,521	609	hypothetical protein	COG2856 Predicted Zn peptidase
IYC_10134	220,695	221,846	1,152	hypothetical protein	COG5279 Uncharacterized protein involved in cytokinesis, contains TGc (transglutaminase/protease-like) domain
IYC_10139	222,171	224,081	1,911	cell wall binding repeat 2-containing protein	COG2247 Putative cell wall-binding domain
IYC_10144	224,188	224,688	501	hypothetical protein	COG3542 Uncharacterized conserved protein
IYC_10149	224,776	225,108	333	hypothetical protein	
IYC_10154	225,437	225,676	240	hypothetical protein	COG2608 Copper chaperone
IYC_10159	225,981	226,742	762	beta-lactamase domain-containing protein	COG0491 Zn-dependent hydrolases, including glyoxylases
IYC_10164	226,865	227,026	162	hypothetical protein	
IYC_10179	228,125	229,978	1,854	glucosyl hydrolase family protein	COG3325 Chitinase
IYC_10184	230,889	231,515	627	hypothetical protein	COG0500 SAM-dependent methyltransferases
IYC_10194	232,206	232,319	114	hypothetical protein	
IYC_10204	232,943	233,395	453	S-ribosylhomocysteinase	COG1854 LuxS protein involved in autoinducer AI2 synthesis
IYC_10209	233,379	234,287	909	cysteine synthase family protein	COG0031 Cysteine synthase
IYC_10214	234,468	235,766	1,299	histidine kinase	COG0642 Signal transduction histidine kinase
IYC_10219	235,763	236,440	678	two component transcriptional regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_10224	236,437	236,991	555	hypothetical protein	COG3963 Phospholipid N-methyltransferase
IYC_10229	237,259	237,900	642	snare associated golgi family protein	
IYC_10234	238,173	238,736	564	NADPH-dependent FMN reductase	COG0655 Multimeric flavodoxin WrbA
IYC_10239	238,738	239,187	450	hypothetical protein	COG1846 Transcriptional regulators
IYC_10244	239,366	239,548	183	3-oxoacyl-[acyl-carrier-protein] synthase III-like protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_10249	239,740	241,293	1,554	ABC transporter, ATP-binding protein	COG0488 ATPase components of ABC transporters with duplicated ATPase domains
IYC_10254	241,492	241,767	276	hypothetical protein	COG2197 Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain
IYC_10259	242,261	242,428	168	hypothetical protein	
IYC_10264	242,710	244,176	1,467	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_10269	244,339	245,337	999	3-oxoacyl-[acyl-carrier-protein] synthase III-like protein	COG0332 3-oxoacyl-[acyl-carrier-protein] synthase III
IYC_10274	245,466	245,726	261	tfoX domain-containing protein	COG3070 Regulator of competence-specific genes
IYC_10279	245,812	246,777	966	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_10284	246,967	247,524	558	hypothetical protein	
IYC_10289	247,905	248,351	447	hypothetical protein	COG1396 Predicted transcriptional regulators
IYC_10294	248,729	249,259	531	hypothetical protein	COG1695 Predicted transcriptional regulators
IYC_10299	249,382	250,779	1,398	drug/sodium antiporter	COG0534 Na+-driven multidrug efflux pump
IYC_10304	250,823	251,011	189	hypothetical protein	
IYC_10319	252,157	253,134	978	putative lipoate-protein ligase A	COG0095 Lipoate-protein ligase A
IYC_10324	253,151	254,539	1,389	glycine cleavage system, dihydrolipoamide dehydrogenase	COG1249 Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes
IYC_10329	254,859	256,313	1,455	glycine dehydrogenase subunit 2	COG1003 Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain
IYC_10334	256,313	257,653	1,341	glycine dehydrogenase subunit 1	COG0403 Glycine cleavage system protein P (pyridoxal-binding), N-terminal domain
IYC_10339	257,656	258,048	393	glycine cleavage system protein H	COG0509 Glycine cleavage system H protein (lipoate-binding)
IYC_10344	258,313	259,425	1,113	glycine cleavage system aminomethyltransferase T	COG0404 Glycine cleavage system T protein (aminomethyltransferase)
IYC_10349	259,589	259,681	93	hypothetical protein	
IYC_10354	260,064	261,509	1,446	hypothetical protein	
IYC_10359	261,598	262,206	609	hypothetical protein	COG4832 Uncharacterized conserved protein
IYC_10364	262,196	263,272	1,077	hypothetical protein	
IYC_10369	263,351	264,166	816	MerR family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_10374	264,469	265,062	594	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_10379	265,043	265,456	414	Cytidine deaminase	COG0295 Cytidine deaminase
IYC_10384	265,447	266,139	693	Zn-dependent hydrolase	COG0491 Zn-dependent hydrolases, including glyoxylases
IYC_10389	266,575	267,060	486	AraC family transcriptional regulator	COG3449 DNA gyrase inhibitor
IYC_10394	267,081	267,446	366	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_10399	267,729	267,887	159	hypothetical protein	COG0534 Na+-driven multidrug efflux pump
IYC_10404	267,912	269,924	2,013	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_10409	270,383	271,621	1,239	hypothetical protein	COG2710 Nitrogenase molybdenum-iron protein, alpha and beta chains
IYC_10414	271,608	272,903	1,296	hypothetical protein	COG2710 Nitrogenase molybdenum-iron protein, alpha and beta chains
IYC_10419	272,906	273,655	750	nitrogenase iron protein	COG1348 Nitrogenase subunit NifH (ATPase)
IYC_10424	273,685	274,470	786	ferrichrome transport ATP-binding protein FhuC	COG1120 ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components
IYC_10429	274,745	275,785	1,041	ferrichrome ABC transporter, permease protein FhuG	COG0609 ABC-type Fe3+-siderophore transport system,permease component
IYC_10434	275,785	276,825	1,041	ferrichrome ABC transporter, permease protein FhuB	COG0609 ABC-type Fe3+-siderophore transport system,permease component
IYC_10439	276,913	277,014	102	hypothetical protein	
IYC_10454	278,264	279,253	990	ABC transporter iron chelate uptake transporter periplasmic-binding protein	COG0614 ABC-type Fe3+-hydroxamate transport system,periplasmic component

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_10459	279,901	280,725	825	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_10464	280,902	281,054	153	hypothetical protein	
IYC_10469	281,186	281,488	303	hypothetical protein	COG3877 Uncharacterized protein conserved in bacteria
IYC_10474	281,678	282,142	465	transcription elongation factor GreA	COG0782 Transcription elongation factor
IYC_10479	282,278	283,237	960	dihydrouridine synthase family protein	COG0042 tRNA-dihydrouridine synthase
IYC_10484	283,546	284,223	678	serine/threonine protein kinase	COG0515 Serine/threonine protein kinase
IYC_10489	284,628	285,689	1,062	NlpC/P60 family protein	COG0791 Cell wall-associated hydrolases (invasion-associated proteins)
IYC_10494	286,065	287,252	1,188	peptidase, M56 family protein	
IYC_10499	287,655	287,765	111	hypothetical protein	
IYC_10514	290,129	290,509	381	penicillinase repressor	COG3682 Predicted transcriptional regulator
IYC_10519	291,079	291,684	606	transposase	COG2801 Transposase and inactivated derivatives
IYC_10524	291,684	291,803	120	transposase	
IYC_10544	294,098	294,484	387	penicillinase repressor	COG3682 Predicted transcriptional regulator
IYC_10549	294,658	295,644	987	metallo-beta-lactamase family protein	COG0491 Zn-dependent hydrolases, including glyoxylases
IYC_10554	296,111	296,314	204	hypothetical protein	
IYC_10559	296,507	297,853	1,347	hypothetical protein	
IYC_10564	297,860	298,327	468	hypothetical protein	COG0034 Glutamine phosphoribosylpyrophosphate amidotransferase
IYC_10569	298,813	299,997	1,185	sodium/glutamate symporter	COG0786 Na+/glutamate symporter
IYC_10574	300,665	301,084	420	hypothetical protein	
IYC_10579	301,151	301,912	762	hypothetical protein	COG2461 Uncharacterized conserved protein
IYC_10584	301,938	302,636	699	putative scdA protein	COG2846 Regulator of cell morphogenesis and NO signaling
IYC_10589	302,655	303,887	1,233	metallo-beta-lactamase family protein	COG0426 Uncharacterized flavoproteins
IYC_10594	304,174	304,884	711	Crp/Fnr family transcriptional regulator	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_10599	305,591	306,409	819	phosphotransferase domain-containing protein	COG0613 Predicted metal-dependent phosphoesterases (PHP family)
IYC_10604	306,421	308,262	1,842	PTS system, beta-glucoside-specific, IIABC component	COG1264 Phosphotransferase system IIB components
IYC_10609	308,467	309,306	840	BglG family transcriptional antiterminator	COG3711 Transcriptional antiterminator
IYC_10614	309,803	310,129	327	CGGC domain-containing protein	COG5561 Predicted metal-binding protein
IYC_10619	310,150	310,908	759	putative hemin transport system ATP-binding protein fecE	COG1120 ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components
IYC_10624	310,915	312,015	1,101	iron(III) dicitrate transport system permease protein fecD	COG0609 ABC-type Fe3+-siderophore transport system,permease component
IYC_10629	312,019	313,077	1,059	putative iron ABC transporter, iron-binding protein	COG0614 ABC-type Fe3+-hydroxamate transport system,periplasmic component
IYC_10634	313,809	314,207	399	putative mutator mutT protein	COG0494 NTP pyrophosphohydrolases including oxidative damage repair enzymes
IYC_10639	314,306	317,272	2,967	DNA repair helicase Rad25	COG3886 Predicted HKD family nuclease
IYC_10644	317,401	318,069	669	putative sanA protein	COG2949 Uncharacterized membrane protein
IYC_10659	318,949	319,824	876	LysR family transcriptional regulator	COG0583 Transcriptional regulator
IYC_10664	319,949	321,139	1,191	hypothetical protein	COG1478 Uncharacterized conserved protein
IYC_10669	321,374	322,063	690	hypothetical protein	
IYC_10674	322,599	324,185	1,587	periplasmic dipeptide transport protein	COG0747 ABC-type dipeptide transport system, periplasmic component
IYC_10679	324,186	325,136	951	nickel transport system permease protein NikB	COG0601 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components
IYC_10684	325,138	325,953	816	glutathione transport system permease protein GsiD	COG1173 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_10689	326,127	327,068	942	dipeptide transport ATP-binding protein DppD	COG0444 ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component
IYC_10694	327,052	327,822	771	glutathione import ATP-binding protein GsiA	COG4608 ABC-type oligopeptide transport system, ATPase component
IYC_10699	328,157	328,846	690	hypothetical protein	
IYC_10704	329,032	330,039	1,008	AraC family regulatory protein	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_10709	330,207	331,031	825	metallo-beta-lactamase family protein	COG1237 Metal-dependent hydrolases of the beta-lactamase superfamily II
IYC_10714	331,332	332,765	1,434	drug resistance transporter, EmrB/QacA family protein	COG0477 Permeases of the major facilitator superfamily
IYC_10719	332,793	333,230	438	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_10724	333,402	333,890	489	putative rubrerythrin	COG1633 Uncharacterized conserved protein
IYC_10729	334,475	335,752	1,278	ATP-dependent RNA helicase RhlE	COG0513 Superfamily II DNA and RNA helicases
IYC_10734	336,271	336,618	348	ferredoxin	COG1145 Ferredoxin
IYC_10739	337,087	337,197	111	hypothetical protein	
IYC_10744	337,333	339,228	1,896	MutS domain-containing protein	COG1193 Mismatch repair ATPase (MutS family)
IYC_10749	339,760	340,146	387	glyoxalase I	COG0346 Lactoylglutathione lyase and related lyases
IYC_10764	341,222	341,551	330	tRNA-binding protein	COG0073 EMAP domain
IYC_10779	343,327	344,505	1,179	cyclopropane-fatty-acyl-phospholipid synthase	COG2230 Cyclopropane fatty acid synthase and related methyltransferases
IYC_10784	344,673	345,968	1,296	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_10789	345,965	346,777	813	response regulator	COG0784 FOG: CheY-like receiver
IYC_10794	346,920	348,047	1,128	hypothetical protein	COG1299 Phosphotransferase system, fructose-specific IIC component
IYC_10799	348,368	349,012	645	AcrR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_10804	349,166	350,833	1,668	MDR-type permease	COG0477 Permeases of the major facilitator superfamily
IYC_10809	351,149	352,252	1,104	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_10814	352,242	352,934	693	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_10819	352,955	353,671	717	D-alanyl-D-alanine carboxypeptidase family protein	COG1876 D-alanyl-D-alanine carboxypeptidase
IYC_10824	353,963	355,384	1,422	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_10829	355,390	356,073	684	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_10834	356,251	357,318	1,068	putative lipoprotein	COG0457 FOG: TPR repeat
IYC_10849	358,860	359,252	393	DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_10854	359,265	360,176	912	CAAX amino terminal protease family protein	COG1266 Predicted metal-dependent membrane protease
IYC_10859	360,608	360,703	96	hypothetical protein	
IYC_10864	361,362	362,063	702	hypothetical protein	
IYC_10869	362,065	362,520	456	LytR family transcriptional regulator	COG3279 Response regulator of the LytR/AlgR family
IYC_10874	362,723	363,259	537	hypothetical protein	
IYC_10879	363,275	363,595	321	PadR family transcriptional regulator	COG1695 Predicted transcriptional regulators
IYC_10884	364,081	364,677	597	ATP-dependent Clp protease proteolytic subunit	COG0740 Protease subunit of ATP-dependent Clp proteases
IYC_10889	364,785	366,176	1,392	asparaginyl-tRNA synthetase	COG0017 Aspartyl/asparaginyl-tRNA synthetases
IYC_10894	367,163	368,386	1,224	NupC family nucleoside transporter	COG1972 Nucleoside permease
IYC_10899	368,714	370,126	1,413	sigma-54 dependent transcriptional regulator	COG3829 Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding domains
IYC_10904	370,495	372,405	1,911	sulfatase family protein	COG1368 Phosphoglycerol transferase and related proteins, alkaline phosphatase superfamily

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_10909	372,835	373,347	513	hypothetical protein	COG4843 Uncharacterized protein conserved in bacteria
IYC_10914	373,662	373,760	99	hypothetical protein	
IYC_10919	373,750	374,616	867	degV family protein	COG1307 Uncharacterized protein conserved in bacteria
IYC_10924	374,809	375,072	264	hypothetical protein	
IYC_10929	375,533	376,951	1,419	xanthine/uracil/vitamin C permease	COG2252 Permeases
IYC_10934	377,041	377,325	285	YCII-like protein	COG2350 Uncharacterized protein conserved in bacteria
IYC_10939	377,430	378,023	594	transcriptional regulator	COG1309 Transcriptional regulator
IYC_10944	378,179	378,643	465	acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_10949	379,234	379,914	681	ABC transporter ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_10954	379,929	382,397	2,469	putative ABC transporter, permease protein	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_10959	382,394	382,873	480	GNAT family acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_10964	382,896	383,645	750	putative lipoprotein	
IYC_10969	383,679	385,391	1,713	putative ABC transporter ATP-binding protein/permease	COG2274 ABC-type bacteriocin/lantibiotic exporters,contain an N-terminal double-glycine peptidase domain
IYC_10974	385,426	385,923	498	hypothetical protein	
IYC_10979	386,143	386,733	591	hypothetical protein	
IYC_10984	386,726	388,078	1,353	radical SAM domain-containing protein	COG0641 Arylsulfatase regulator (Fe-S oxidoreductase)
IYC_10989	388,685	388,831	147	hypothetical protein	
IYC_10994	388,979	389,083	105	hypothetical protein	
IYC_10999	389,533	389,658	126	hypothetical protein	
IYC_11004	389,699	389,818	120	hypothetical protein	
IYC_11009	390,412	391,113	702	LytTr family DNA-binding response regulator	COG3279 Response regulator of the LytR/AlgR family
IYC_11014	392,051	392,659	609	signal peptidase I	COG0681 Signal peptidase I
IYC_11019	392,711	393,148	438	hypothetical protein	
IYC_11024	393,138	393,350	213	DNA binding protein	COG1476 Predicted transcriptional regulators
IYC_11029	393,492	393,842	351	tRNA-binding domain protein	COG0073 EMAP domain
IYC_11034	394,108	394,959	852	pentapeptide repeat-containing protein	COG1357 Uncharacterized low-complexity proteins
IYC_11039	395,282	396,286	1,005	CAAX amino terminal protease family protein	COG1266 Predicted metal-dependent membrane protease
IYC_11044	396,549	397,382	834	hypothetical protein	COG1633 Uncharacterized conserved protein
IYC_11049	397,465	398,184	720	rubrerythrin	COG1633 Uncharacterized conserved protein
IYC_11054	398,308	399,630	1,323	proton/sodium-glutamate symporter	COG1301 Na+/H+-dicarboxylate symporters
IYC_11059	399,870	400,739	870	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_11064	401,086	402,927	1,842	putative spermidine/putrescine ABC transporter protein	COG1177 ABC-type spermidine/putrescine transport system, permease component II; putative spermidine/putrescine ABC transporter, permease protein/spermidine/putrescine-binding protein
IYC_11069	402,930	403,892	963	spermidine/putrescine ABC transporter, permease protein PotB	COG1176 ABC-type spermidine/putrescine transport system, permease component I
IYC_11074	403,892	404,974	1,083	spermidine/putrescine ABC transporter, ATP-binding protein	COG3842 ABC-type spermidine/putrescine transport systems, ATPase components
IYC_11079	405,640	407,337	1,698	putative methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_11084	407,695	408,456	762	putative ABC transporter, permease protein	
IYC_11089	408,456	409,151	696	ABC transporter ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_11094	409,144	409,524	381	GntR family transcriptional regulator	COG1725 Predicted transcriptional regulators
IYC_11099	409,852	410,973	1,122	SecC motif-containing protein	COG0653 Preprotein translocase subunit SecA (ATPase, RNA helicase)

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_11104	411,194	411,316	123	hypothetical protein	
IYC_11109	411,576	411,752	177	hypothetical protein	
IYC_11114	411,885	412,568	684	putative proton-coupled thiamine transporter YuaJ	COG3859 Predicted membrane protein
IYC_11119	412,805	413,203	399	large conductance mechanosensitive channel protein	COG1970 Large-conductance mechanosensitive channel
IYC_11124	413,679	415,337	1,659	ABC transporter ATP-binding protein	COG0488 ATPase components of ABC transporters with duplicated ATPase domains
IYC_11129	415,808	417,799	1,992	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_11134	418,633	419,322	690	hypothetical protein	
IYC_11139	419,738	421,381	1,644	carbon starvation protein CstA	COG1966 Carbon starvation protein, predicted membrane protein
IYC_11144	421,393	421,902	510	hypothetical protein	
IYC_11149	422,111	423,019	909	hypothetical protein	COG1073 Hydrolases of the alpha/beta superfamily
IYC_11154	423,345	423,557	213	[2Fe-2S]-binding domain-containing protein	COG0578 Glycerol-3-phosphate dehydrogenase
IYC_11159	423,909	424,358	450	hypothetical protein	
IYC_11164	424,492	424,944	453	peptide deformylase	COG0242 N-formylmethionyl-tRNA deformylase
IYC_11169	424,945	425,181	237	putative DNA-binding protein	COG1476 Predicted transcriptional regulators
Genomic Scaffold: NZ_JH470523					
IYC_11174	626	1,627	1,002	L-asparaginase, type II	COG0252 L-asparaginase/archaeal Glu-tRNAGln amidotransferase subunit D
IYC_11179	1,821	2,699	879	hypothetical protein	COG1561 Uncharacterized stress-induced protein
IYC_11184	2,715	2,990	276	hypothetical protein	COG2052 Uncharacterized protein conserved in bacteria
IYC_11189	2,990	3,619	630	guanylate kinase	COG0194 Guanylate kinase
IYC_11194	3,600	3,818	219	DNA-directed RNA polymerase subunit omega	COG1758 DNA-directed RNA polymerase, subunit K/omega
IYC_11199	3,820	5,007	1,188	bifunctional phosphopantothenoylcysteine decarboxylase/phosphopantothenate synthase	COG0452 Phosphopantothenoylcysteine synthetase/decarboxylase
IYC_11204	5,380	7,587	2,208	primosome assembly protein PriA	COG1198 Primosomal protein N' (replication factor Y) - superfamily II helicase
IYC_11209	7,613	8,056	444	peptide deformylase	COG0242 N-formylmethionyl-tRNA deformylase
IYC_11214	8,073	9,014	942	methionyl-tRNA formyltransferase	COG0223 Methionyl-tRNA formyltransferase
IYC_11219	9,316	10,032	717	putative neutral zinc metallopeptidase	COG2738 Predicted Zn-dependent protease
IYC_11224	10,053	11,381	1,329	16S rRNA methyltransferase B	COG0144 tRNA and rRNA cytosine-C5-methylases
IYC_11229	11,393	12,421	1,029	ribosomal RNA large subunit methyltransferase N	COG0820 Predicted Fe-S-cluster redox enzyme
IYC_11234	12,440	13,189	750	protein phosphatase family protein	COG0631 Serine/threonine protein phosphatase
IYC_11239	13,167	15,029	1,863	protein kinase	COG0515 Serine/threonine protein kinase
IYC_11244	15,164	16,042	879	GTPase RsgA	COG1162 Predicted GTPases
IYC_11249	16,036	16,686	651	ribulose-phosphate 3-epimerase	COG0036 Pentose-5-phosphate-3-epimerase
IYC_11254	16,688	17,332	645	thiamine pyrophosphokinase	COG1564 Thiamine pyrophosphokinase
IYC_11259	17,704	17,859	156	hypothetical protein	
IYC_11264	18,107	18,298	192	50S ribosomal protein L28	COG0227 Ribosomal protein L28
IYC_11269	18,496	18,846	351	hypothetical protein	COG1302 Uncharacterized protein conserved in bacteria
IYC_11274	18,861	20,486	1,626	DAK2 domain-containing protein	COG1461 Predicted kinase related to dihydroxyacetone kinase
IYC_11279	20,679	22,718	2,040	ATP-dependent DNA helicase RecG	COG1200 RecG-like helicase
IYC_11284	22,978	23,535	558	putative methyltransferase	COG0742 N6-adenine-specific methylase
IYC_11289	23,537	24,031	495	phosphopantetheine adenylyltransferase	COG0669 Phosphopantetheine adenylyltransferase
IYC_11294	24,044	24,568	525	hypothetical protein	COG3599 Cell division initiation protein

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_11299	24,569	25,735	1,167	nucleoside recognition domain-containing protein	COG3314 Uncharacterized protein conserved in bacteria
IYC_11304	25,795	27,024	1,230	hypothetical protein	COG1323 Predicted nucleotidyltransferase
IYC_11309	27,358	29,112	1,755	sensory box sigma-54 dependent transcriptional regulator	COG0517 FOG: CBS domain
IYC_11314	29,287	30,597	1,311	putative electron transport protein	COG4656 Predicted NADH:ubiquinone oxidoreductase, subunit RnfC
IYC_11319	30,628	31,107	480	PrdA domain-containing protein	
IYC_11324	31,129	31,299	171	hypothetical protein	
IYC_11329	31,334	31,486	153	D-proline reductase, PrdA proprotein	
IYC_11334	31,775	31,951	177	transposase	COG3666 Transposase and inactivated derivatives
IYC_11339	32,167	32,646	480	hypothetical protein	
IYC_11344	32,651	33,178	528	hypothetical protein	
IYC_11349	33,554	34,858	1,305	putative electron transport protein	COG4656 Predicted NADH:ubiquinone oxidoreductase, subunit RnfC
IYC_11354	34,898	35,560	663	hypothetical protein	
IYC_11359	36,199	>36883	>685	putative electron transport protein	COG4656 Predicted NADH:ubiquinone oxidoreductase, subunit RnfC
Genomic Scaffold: NZ_JH470524					
IYC_11364	399	677	279	hypothetical protein	
IYC_11369	929	1,939	1,011	hypothetical protein	COG2855 Predicted membrane protein
IYC_11374	2,259	3,581	1,323	aminotransferase, class V	COG0520 Selenocysteine lyase
IYC_11379	3,698	4,264	567	HD domain-containing protein	COG2316 Predicted hydrolase (HD superfamily)
IYC_11384	4,402	5,742	1,341	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_11389	6,254	8,932	2,679	putative diguanylate cyclase	COG2199 FOG: GGDEF domain
IYC_11394	9,061	9,735	675	DedA family protein	COG0398 Uncharacterized conserved protein
IYC_11399	9,761	10,327	567	hypothetical protein	COG0500 SAM-dependent methyltransferases
IYC_11404	10,371	11,864	1,494	putative lipoprotein	
IYC_11409	11,932	13,326	1,395	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_11414	13,336	14,022	687	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_11419	14,129	14,941	813	polysaccharide deacetylase family protein	COG0726 Predicted xylanase/chitin deacetylase
IYC_11424	15,304	17,856	2,553	hypothetical protein	COG1511 Predicted membrane protein
IYC_11429	17,888	19,984	2,097	RND superfamily exporter	COG1033 Predicted exporters of the RND superfamily
IYC_11434	20,336	20,863	528	PAP2 family protein	COG0671 Membrane-associated phospholipid phosphatase
IYC_11439	20,987	21,997	1,011	hypothetical protein	
IYC_11444	22,348	22,827	480	hypothetical protein	COG2188 Transcriptional regulators
IYC_11449	23,012	23,605	594	hypothetical protein	
IYC_11454	23,664	23,951	288	hypothetical protein	COG4378 Uncharacterized protein conserved in bacteria
IYC_11459	24,267	26,855	2,589	bifunctional acetaldehyde-CoA/alcohol dehydrogenase	COG1012 NAD-dependent aldehyde dehydrogenases
IYC_11464	27,382	28,464	1,083	metallopeptidase, family M24	COG0006 Xaa-Pro aminopeptidase
IYC_11469	28,605	30,005	1,401	amino acid permease family protein	COG0531 Amino acid transporters
IYC_11474	30,364	30,909	546	DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_11479	31,035	31,901	867	PP-loop family protein	COG0037 Predicted ATPase of the PP-loop superfamilyimplicated in cell cycle control
IYC_11484	32,134	33,987	1,854	sensory box histidine kinase	COG0642 Signal transduction histidine kinase
IYC_11489	34,071	34,205	135	hypothetical protein	
IYC_11494	34,289	34,918	630	accessory gene regulator B	COG4512 Membrane protein putatively involved in post-translational modification of the autoinducing quorum-sensing peptide
IYC_11499	34,905	35,372	468	hypothetical protein	
IYC_11504	35,601	37,457	1,857	sensory box histidine kinase	COG0642 Signal transduction histidine kinase
IYC_11509	37,469	37,936	468	hypothetical protein	
IYC_11514	38,327	38,836	510	GNAT family acetyltransferase	COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins
IYC_11519	39,067	40,770	1,704	GGDEF/HD domain-containing protein	COG2199 FOG: GGDEF domain

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_11524	40,836	40,970	135	putative autoinducer prepeptide	
IYC_11529	40,954	41,643	690	putative accessory gene regulator protein B	
IYC_11534	41,636	42,103	468	hypothetical protein	
IYC_11539	42,282	42,833	552	hypothetical protein	
IYC_11544	43,115	44,623	1,509	hypothetical protein	COG4868 Uncharacterized protein conserved in bacteria
IYC_11549	44,757	45,650	894	radical SAM domain-containing protein	COG1533 DNA repair photolyase
IYC_11554	45,959	46,609	651	putative lipoprotein	
IYC_11559	46,908	47,603	696	lrgB family protein	COG1346 Putative effector of murein hydrolase
IYC_11564	47,603	47,962	360	lrgA family protein	COG1380 Putative effector of murein hydrolase LrgA
IYC_11569	48,528	50,390	1,863	peptide ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_11574	50,390	51,136	747	peptide ABC transporter ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_11579	51,151	52,008	858	undecaprenyl pyrophosphate phosphatase	COG1968 Uncharacterized bacitracin resistance protein
IYC_11584	52,185	53,165	981	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_11589	53,178	53,870	693	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_11594	54,331	54,903	573	xanthine phosphoribosyltransferase	COG0503 Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins
IYC_11599	54,926	56,227	1,302	xanthine/uracil permease family protein	COG2233 Xanthine/uracil permeases
IYC_11604	56,707	57,087	381	putative iron-dependent repressor	COG1321 Mn-dependent transcriptional regulator
IYC_11609	57,142	58,866	1,725	ABC transporter, ATP-binding/permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_11614	58,863	59,693	831	ferrichrome ABC transporter, ATP-binding protein	COG1120 ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components
IYC_11619	59,704	60,717	1,014	iron compound ABC transporter permease	COG0609 ABC-type Fe3+-siderophore transport system,permease component
IYC_11624	60,714	61,742	1,029	iron compound ABC transporter, permease protein	COG0609 ABC-type Fe3+-siderophore transport system,permease component
IYC_11629	61,769	62,698	930	iron compound ABC transporter, iron compound-binding protein	COG0614 ABC-type Fe3+-hydroxamate transport system,periplasmic component
IYC_11634	63,009	63,977	969	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_11639	64,122	65,450	1,329	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_11644	65,484	66,254	771	iron chelate ABC transporter ATP-binding protein	COG1120 ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components
IYC_11649	66,256	67,374	1,119	iron chelate ABC transporter permease protein	COG0609 ABC-type Fe3+-siderophore transport system,permease component
IYC_11654	67,377	68,414	1,038	iron chelate ABC transporter solute-binding protein	COG0614 ABC-type Fe3+-hydroxamate transport system,periplasmic component
IYC_11659	68,662	69,699	1,038	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_11664	70,088	70,843	756	hypothetical protein	
IYC_11669	70,961	72,301	1,341	putative membrane-associated ATP-binding domain-containing protein	COG2972 Predicted signal transduction protein with a C-terminal ATPase domain
IYC_11674	72,305	73,027	723	response regulator	COG3279 Response regulator of the LytR/AlgR family
IYC_11679	73,391	75,700	2,310	dolichyl-phosphate-mannose-protein mannosyltransferase family protein	COG1807 4-amino-4-deoxy-L-arabinose transferase andrelated glycosyltransferases of PMT family
IYC_11684	75,827	76,246	420	putative cell wall teichoic acid glycosylation protein GtcA	COG2246 Predicted membrane protein

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_11689	76,230	77,198	969	glycosyl transferase, group 2 family protein	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_11694	77,450	78,838	1,389	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_11699	78,835	79,509	675	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_11704	79,916	81,076	1,161	arsenite-activated ATPase ArsA	COG0003 Oxyanion-translocating ATPase
IYC_11709	81,093	81,317	225	hypothetical protein	
IYC_11714	81,329	82,495	1,167	arsenite-activated ATPase ArsA	COG0003 Oxyanion-translocating ATPase
IYC_11719	82,576	83,076	501	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_11724	83,275	85,074	1,800	adenine deaminase	COG1001 Adenine deaminase
IYC_11729	85,329	86,333	1,005	putative 2-dehydro-3-deoxygluconokinase	COG0524 Sugar kinases, ribokinase family
IYC_11734	86,335	86,982	648	2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase	COG0800 2-keto-3-deoxy-6-phosphogluconate aldolase
IYC_11739	87,034	88,749	1,716	dihydroxy-acid dehydratase	COG0129 Dihydroxyacid dehydratase/phosphogluconate dehydratase
IYC_11744	88,887	89,360	474	PTS system, glucose/maltose-specific, IIA component	COG2190 Phosphotransferase system IIA components
IYC_11749	89,373	90,707	1,335	maltose-6'-phosphate glucosidase	COG1486 Alpha-galactosidases/6-phospho-beta-glucosidases, family 4 of glycosyl hydrolases
IYC_11754	90,855	91,034	180	hypothetical protein	
IYC_11759	91,175	92,758	1,584	PTS system glucose/maltose-specific transporter subunit IIBC	COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific
IYC_11764	92,877	93,638	762	RpiR family transcriptional regulator	COG1737 Transcriptional regulators
IYC_11769	93,866	95,362	1,497	methyl-accepting chemotaxis sensory transducer	COG0840 Methyl-accepting chemotaxis protein
IYC_11774	95,727	96,365	639	hypothetical protein	COG2910 Putative NADH-flavin reductase
IYC_11779	96,392	96,997	606	2-hydroxychromene-2-carboxylate isomerase familyprotein	COG2761 Predicted dithiol-disulfide isomerase involved in polyketide biosynthesis
IYC_11784	97,150	97,581	432	rrf2 family protein	COG1959 Predicted transcriptional regulator
IYC_11789	97,732	98,286	555	hypothetical protein	COG4420 Predicted membrane protein
IYC_11794	98,628	98,825	198	cold shock protein	COG1278 Cold shock proteins
IYC_11799	99,083	99,844	762	SIS domain-containing protein	COG1737 Transcriptional regulators
IYC_11804	100,185	100,694	510	PTS system glucose/glucoside family transporter subunit IIA	COG2190 Phosphotransferase system IIA components
IYC_11809	100,952	102,277	1,326	maltose-6'-phosphate glucosidase	COG1486 Alpha-galactosidases/6-phospho-beta-glucosidases, family 4 of glycosyl hydrolases
IYC_11814	102,594	104,165	1,572	PTS system glucose/glucoside family transporter subunit IIBC	COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific
IYC_11819	104,629	105,843	1,215	alanine racemase	COG0787 Alanine racemase
IYC_11824	106,074	108,044	1,971	peptide ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_11829	108,058	108,828	771	ABC transporter ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_11834	108,948	109,997	1,050	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_11839	110,007	110,705	699	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_11844	110,899	112,377	1,479	spore germination protein GerA	
IYC_11849	112,370	113,461	1,092	hypothetical protein	
IYC_11854	113,448	114,569	1,122	germination protein, Ger(x)C family	COG0245 2C-methyl-D-erythritol 2,4-cyclodiphosphatesynthase
IYC_11859	114,716	114,943	228	DNA polymerase III subunit epsilon	COG0847 DNA polymerase III, epsilon subunit and related 3'-5' exonucleases
IYC_11864	115,169	116,260	1,092	sensory transduction histidine kinase	COG0642 Signal transduction histidine kinase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_11869	116,250	116,945	696	regulatory protein VanR	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_11874	117,014	118,975	1,962	putative beta-lactamase	COG1680 Beta-lactamase class C and other penicillinbinding proteins
Genomic Scaffold: NZ_JH470525					
IYC_11879	53	>214	>162	transposase IS3/IS911 family protein	COG2963 Transposase and inactivated derivatives
Genomic Scaffold: NZ_JH470526					
IYC_11884	261	2,183	1,923	sensory box sigma-54 dependent transcriptional regulator	COG3284 Transcriptional activator of acetoin/glycerol metabolism
IYC_11889	2,488	3,120	633	hypothetical protein	
IYC_11894	3,153	3,707	555	hypothetical protein	COG3945 Uncharacterized conserved protein
IYC_11899	4,156	5,325	1,170	alanine racemase	COG0787 Alanine racemase
IYC_11904	5,592	7,064	1,473	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_11909	7,288	8,493	1,206	xanthine/uracil permease family protein	COG2252 Permeases
IYC_11914	8,721	9,554	834	hypothetical protein	COG5523 Predicted integral membrane protein
IYC_11919	9,761	10,645	885	short chain dehydrogenase/reductase family oxidoreductase	COG1028 Dehydrogenases with different specificities(related to short-chain alcohol dehydrogenases)
IYC_11924	10,985	12,991	2,007	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_11929	13,216	15,168	1,953	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_11934	15,641	15,802	162	hypothetical protein	
IYC_11939	16,502	17,515	1,014	KWG repeat-containing protein	
IYC_11944	17,669	18,220	552	hypothetical protein	
IYC_11949	18,582	21,815	3,234	helicase, SNF2/RAD54 family protein	COG0553 Superfamily II DNA/RNA helicases, SNF2 family
IYC_11954	21,787	22,059	273	ribokinase	COG0524 Sugar kinases, ribokinase family
IYC_11959	22,125	22,529	405	hypothetical protein	COG1720 Uncharacterized conserved protein
IYC_11964	22,760	22,936	177	hypothetical protein	
IYC_11969	23,124	23,648	525	formylmethanofuran dehydrogenase E subunit-related protein	COG2191 Formylmethanofuran dehydrogenase subunit E
IYC_11974	23,791	25,386	1,596	peptide/opine/nickel uptake ABC transporter substrate-binding protein	COG0747 ABC-type dipeptide transport system, periplasmic component
IYC_11979	25,401	26,375	975	peptide/opine/nickel uptake ABC transporter permease	COG0601 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components
IYC_11984	26,379	27,200	822	peptide/opine/nickel uptake ABC transporter permease	COG1173 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components
IYC_11989	27,231	28,913	1,683	ABC transporter, ATP-binding protein	COG1123 ATPase components of various ABC-type transport systems, contain duplicated ATPase
IYC_11994	29,223	29,810	588	putative DNA polymerase III, epsilon subunit	COG0847 DNA polymerase III, epsilon subunit and related 3'-5' exonucleases
IYC_11999	30,101	31,051	951	hypothetical protein	COG4086 Predicted secreted protein
IYC_12004	31,183	31,443	261	transcriptional regulator, ArsR family protein	COG0640 Predicted transcriptional regulators
IYC_12009	31,643	32,278	636	UbiE/COQ5 family methyltransferase	COG0500 SAM-dependent methyltransferases
IYC_12014	32,450	32,890	441	heat shock protein	COG0071 Molecular chaperone (small heat shock protein)
IYC_12019	33,206	34,405	1,200	major facilitator family protein	COG0477 Permeases of the major facilitator superfamily
IYC_12024	34,549	35,103	555	hypothetical protein	COG1309 Transcriptional regulator
IYC_12029	35,373	36,122	750	DeoR family transcriptional regulator	COG1349 Transcriptional regulators of sugar metabolism
IYC_12034	36,122	37,027	906	fructose-1-phosphate kinase	COG1105 Fructose-1-phosphate kinase and related fructose-6-phosphate kinase (PfkB)

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_12039	37,060	38,931	1,872	PTS system, fructose-specific component family protein	COG1762 Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)
IYC_12044	38,970	40,349	1,380	amino acid/peptide transporter	COG3104 Dipeptide/tripeptide permease
IYC_12049	41,070	42,044	975	hypothetical protein	
IYC_12054	42,226	42,759	534	hypothetical protein	
IYC_12059	43,025	43,180	156	hypothetical protein	
IYC_12064	43,199	43,789	591	PfkB family kinase	COG0524 Sugar kinases, ribokinase family
IYC_12069	43,940	44,575	636	spore coat assembly protein SafA	COG1388 FOG: LysM repeat
IYC_12074	44,882	46,576	1,695	oligoendopeptidase F	COG1164 Oligoendopeptidase F
IYC_12079	46,626	47,036	411	hypothetical protein	
IYC_12084	47,312	49,087	1,776	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_12089	49,297	49,506	210	cold shock protein	COG1278 Cold shock proteins
IYC_12094	49,879	51,882	2,004	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_12099	52,117	53,697	1,581	putative drug resistance ABC transporter, ATP-binding protein	COG0488 ATPase components of ABC transporters with duplicated ATPase domains
IYC_12104	54,039	54,503	465	YbaK/prolyl-tRNA synthetase domain-containing protein	COG2606 Uncharacterized conserved protein
IYC_12109	55,156	56,019	864	sulfite/nitrite reductase family protein	COG2221 Dissimilatory sulfite reductase (desulfoviridin), alpha and beta subunits
IYC_12114	56,092	56,985	894	chemotaxis protein CheV	COG0835 Chemotaxis signal transduction protein
IYC_12119	57,359	58,624	1,266	DEAD-box ATP dependent DNA helicase	COG0513 Superfamily II DNA and RNA helicases
IYC_12124	58,945	59,781	837	RNA-binding protein	COG2996 Uncharacterized protein conserved in bacteria
IYC_12129	59,960	61,405	1,446	ATP-dependent RNA helicase DeaD	COG0513 Superfamily II DNA and RNA helicases
IYC_12134	61,570	66,000	4,431	hypothetical protein	
IYC_12139	66,353	70,432	4,080	leucine-rich repeat-containing protein	COG4886 Leucine-rich repeat (LRR) protein
IYC_12144	70,847	71,782	936	mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase domain-containing protein	COG1705 Muramidase (flagellum-specific)
IYC_12149	71,902	72,417	516	ferritin family protein	COG1528 Ferritin-like protein
IYC_12154	72,834	73,688	855	chemotaxis response regulator/CheC-like domain-containing protein	COG0784 FOG: CheY-like receiver
IYC_12159	73,688	74,149	462	chemotaxis protein CheX	COG1406 Predicted inhibitor of MCP methylation, homolog of CheC
IYC_12164	74,398	75,759	1,362	hypothetical protein	COG3314 Uncharacterized protein conserved in bacteria
IYC_12169	76,498	76,701	204	small, acid-soluble spore protein alpha	
IYC_12174	76,761	76,877	117	small, acid-soluble spore protein beta	
IYC_12179	77,226	77,327	102	hypothetical protein	
IYC_12184	77,246	77,356	111	hesB family protein	
IYC_12189	78,283	79,263	981	site-specific tyrosine recombinase XerC	COG0582 Integrase
IYC_12194	79,495	79,968	474	hypothetical protein	
IYC_12199	80,270	80,875	606	LexA repressor	COG1974 SOS-response transcriptional repressors (RecA-mediated autopeptidases)
IYC_12204	81,071	82,354	1,284	aluminum resistance protein	COG4100 Cystathionine beta-lyase family protein involved in aluminum resistance
IYC_12209	82,455	82,712	258	RNA chaperone Hfq	COG1923 Uncharacterized host factor I protein
IYC_12214	82,799	83,734	936	tRNA delta(2)-isopentenylpyrophosphate transferase	COG0324 tRNA delta(2)-isopentenylpyrophosphate transferase
IYC_12219	83,750	85,744	1,995	DNA mismatch repair protein	COG0323 DNA mismatch repair enzyme (predicted ATPase)
IYC_12224	85,887	88,685	2,799	DNA mismatch repair protein MutS	COG0249 Mismatch repair ATPase (MutS family)

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_12229	88,764	90,116	1,353	(dimethylallyl)adenosine tRNA methylthiotransferase	COG0621 2-methylthioadenine synthetase
IYC_12234	90,591	91,781	1,191	aminotransferase, classes I and II	COG1167 Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs
IYC_12239	91,867	92,028	162	hypothetical protein	
IYC_12244	92,082	93,272	1,191	aspartate aminotransferase	COG0436 Aspartate/tyrosine/aromatic aminotransferase
IYC_12249	93,291	94,169	879	acetyltransferase	COG0456 Acetyltransferases
IYC_12254	94,247	95,305	1,059	putative spore coat protein	
IYC_12259	95,498	95,872	375	S-adenosylmethionine decarboxylase proenzyme	COG1586 S-adenosylmethionine decarboxylase
IYC_12264	96,097	98,013	1,917	4-hydroxy-3-methylbut-2-enyl diphosphate reductase/S1 RNA-binding domain protein	COG0761 Penicillin tolerance protein
IYC_12269	98,076	98,729	654	cytidylate kinase	COG0283 Cytidylate kinase
IYC_12274	98,806	99,177	372	chorismate mutase	COG4401 Chorismate mutase
IYC_12279	99,177	100,400	1,224	pyridine nucleotide-disulfide oxidoreductase family protein	COG2081 Predicted flavoproteins
IYC_12284	100,728	101,993	1,266	glutamate dehydrogenase, NAD-specific	COG0334 Glutamate dehydrogenase/leucine dehydrogenase
IYC_12289	102,227	103,111	885	RpiR family transcriptional regulator	COG1737 Transcriptional regulators
IYC_12294	103,510	104,220	711	ribosomal large subunit pseudouridine synthase B	COG1187 16S rRNA uridine-516 pseudouridylate synthase and related pseudouridylate synthases
IYC_12299	104,320	105,609	1,290	hypothetical protein	COG3875 Uncharacterized conserved protein
IYC_12304	105,894	107,174	1,281	hypothetical protein	COG3681 Uncharacterized conserved protein
IYC_12309	107,474	108,877	1,404	group 2 family glycosyl transferase protein	COG1215 Glycosyltransferases, probably involved in cell wall biogenesis
IYC_12314	109,248	109,439	192	hypothetical protein	
IYC_12319	109,822	111,843	2,022	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_12324	111,885	112,571	687	putative cytochrome c-type biogenesis protein	COG4232 Thiol:disulfide interchange protein
IYC_12329	112,573	113,004	432	thioredoxin family protein	COG0526 Thiol-disulfide isomerase and thioredoxins
IYC_12334	113,119	113,955	837	carbon-nitrogen family hydrolase	COG0388 Predicted amidohydrolase
IYC_12339	114,069	114,626	558	hypothetical protein	
IYC_12344	115,034	116,467	1,434	monovalent cation:proton antiporter-2 (CPA2) family protein	COG0475 Kef-type K+ transport systems, membrane components
IYC_12349	116,603	116,947	345	4-carboxymuconolactone decarboxylase	COG0599 Uncharacterized homolog of gamma-carboxymuconolactone decarboxylase subunit
IYC_12354	117,229	117,729	501	acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_12359	117,929	118,435	507	putative lipoprotein	
IYC_12364	118,779	118,979	201	hypothetical protein	COG0393 Uncharacterized conserved protein
IYC_12369	118,960	119,193	234	hypothetical protein	COG0573 ABC-type phosphate transport system, permease component
IYC_12374	119,327	119,575	249	hypothetical protein	COG1905 NADH:ubiquinone oxidoreductase 24 kD subunit
IYC_12379	119,572	121,296	1,725	sensory box-containing [Fe] hydrogenase	COG4624 Iron only hydrogenase large subunit, C-terminal domain
IYC_12384	121,293	122,462	1,170	sporulation domain-containing protein	COG2208 Serine phosphatase RsbU, regulator of sigmasubunit
IYC_12389	122,593	124,332	1,740	sensory box-containing [Fe] hydrogenase	COG4624 Iron only hydrogenase large subunit, C-terminal domain
IYC_12394	124,449	125,453	1,005	hydrogenase expression/formation protein HypE	COG0309 Hydrogenase maturation factor
IYC_12399	125,593	126,666	1,074	hydrogenase isoenzymes formation protein HypD	COG0409 Hydrogenase maturation factor
IYC_12404	126,668	129,256	2,589	[NiFe] hydrogenase maturation protein HypF	COG0068 Hydrogenase maturation factor
IYC_12409	129,270	129,530	261	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_12414	129,756	130,265	510	heme/steroid binding domain-containing protein	COG4892 Predicted heme/steroid binding protein
IYC_12419	130,365	130,796	432	hydrogenase maturation protease	COG0680 Ni,Fe-hydrogenase maturation factor
IYC_12424	130,829	132,244	1,416	[Ni/Fe] hydrogenase, large subunit	COG0374 Ni,Fe-hydrogenase I large subunit
IYC_12429	132,345	133,217	873	[Ni/Fe] hydrogenase, small subunit	COG1740 Ni,Fe-hydrogenase I small subunit
IYC_12434	133,492	134,193	702	TraX family protein	COG0477 Permeases of the major facilitator superfamily
IYC_12439	134,443	135,816	1,374	mate efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_12444	136,498	136,665	168	hypothetical protein	
IYC_12449	137,323	137,502	180	carbamoyl phosphate synthase large subunit	COG0458 Carbamoylphosphate synthase large subunit (split gene in MJ)
IYC_12454	137,867	138,034	168	hypothetical protein	
IYC_12459	138,558	139,418	861	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_12464	139,445	139,903	459	transcription activator effector binding protein	COG4978 Transcriptional regulator, effector-bindingdomain/component
IYC_12469	139,905	140,858	954	ABC transporter ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_12474	140,861	141,610	750	ABC transporter permease	COG0842 ABC-type multidrug transport system, permease component
IYC_12479	141,734	143,290	1,557	drug resistance ABC transporter RA1 family, ATP-binding protein	COG0488 ATPase components of ABC transporters with duplicated ATPase domains
IYC_12484	143,652	145,385	1,734	[Fe] hydrogenase	COG3383 Uncharacterized anaerobic dehydrogenase
IYC_12489	145,405	147,300	1,896	putative [Fe] hydrogenase, electron-transfer subunit	COG1894 NADH:ubiquinone oxidoreductase, NADH-binding (51 kD) subunit
IYC_12494	147,312	147,791	480	putative [Fe] hydrogenase, electron-transfer subunit	COG1905 NADH:ubiquinone oxidoreductase 24 kD subunit
IYC_12499	148,268	149,455	1,188	monovalent cation:proton antiporter-2 (CPA2) family protein	COG0025 NhaP-type Na+/H+ and K+/H+ antiporters
IYC_12504	149,718	150,596	879	ADP-ribosylglycohydrolase family protein	COG1397 ADP-ribosylglycohydrolase
IYC_12509	150,835	151,929	1,095	hypothetical protein	COG0628 Predicted permease
IYC_12514	152,033	153,169	1,137	D-alanyl-D-alanine carboxypeptidase	COG1686 D-alanyl-D-alanine carboxypeptidase
IYC_12519	153,353	153,874	522	hypothetical protein	
IYC_12524	153,942	154,382	441	sporulation protein YtfJ	COG3874 Uncharacterized conserved protein
IYC_12529	154,444	155,025	582	segregation and condensation protein B	COG1386 Predicted transcriptional regulator containing the HTH domain
IYC_12534	155,018	155,764	747	segregation and condensation protein A	COG1354 Uncharacterized conserved protein
IYC_12539	155,975	156,100	126	hypothetical protein	
IYC_12544	156,169	157,392	1,224	D-alanyl-D-alanine carboxypeptidase	COG1686 D-alanyl-D-alanine carboxypeptidase
IYC_12549	157,525	158,247	723	hypothetical protein	
IYC_12554	158,406	159,464	1,059	endonuclease/exonuclease/phosphatase family protein	
IYC_12559	159,616	159,864	249	hypothetical protein	
IYC_12564	159,966	161,270	1,305	pyrimidine-nucleoside phosphorylase	COG0213 Thymidine phosphorylase
IYC_12569	161,549	162,364	816	purine nucleoside phosphorylase	COG0005 Purine nucleoside phosphorylase
IYC_12574	162,402	163,277	876	tyrosine recombinase XerD	COG4974 Site-specific recombinase XerD
IYC_12579	163,329	163,556	228	hypothetical protein	
IYC_12584	163,547	164,200	654	stage II sporulation protein M	COG1300 Uncharacterized membrane protein
IYC_12589	164,418	164,954	537	NUDIX family hydrolase	COG0494 NTP pyrophosphohydrolases including oxidative damage repair enzymes
IYC_12594	165,216	166,742	1,527	inner membrane transporter YjeM	COG0531 Amino acid transporters
IYC_12599	167,593	>167956	>364	UDP-N-acetylglucosamine 2-epimerase	COG0381 UDP-N-acetylglucosamine 2-epimerase
Genomic Scaffold: NZ_JH470527					
IYC_12604	<1	>474	>474	N-acetylmuramoyl-L-alanine amidase	

locus_tag	Minimum	Maximum	Length	Product	COG Group
Genomic Scaffold: NZ_JH470528					
IYC_12609	437	1,459	1,023	zinc-binding dehydrogenase family oxidoreductase	COG1063 Threonine dehydrogenase and related Zn-dependent dehydrogenases
IYC_12614	1,544	2,401	858	AP endonuclease	COG1082 Sugar phosphate isomerases/epimerases
IYC_12619	2,484	3,623	1,140	putative 3-dehydroquinate synthase	COG0337 3-dehydroquinate synthetase
IYC_12624	4,065	5,084	1,020	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_12629	5,081	5,752	672	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_12634	5,950	6,660	711	hypothetical protein	
IYC_12639	6,653	8,812	2,160	hypothetical protein	COG0308 Aminopeptidase N
IYC_12644	8,884	9,747	864	ABC transporter ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_12649	10,220	10,417	198	thiamine biosynthesis protein ThiS	COG2104 Sulfur transfer protein involved in thiamine biosynthesis
IYC_12654	10,419	11,228	810	thiamine biosynthesis protein ThiF	COG0476 Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2
IYC_12659	11,366	12,154	789	thiazole synthase	COG2022 Uncharacterized enzyme of thiazole biosynthesis
IYC_12664	12,168	13,325	1,158	thiamine biosynthesis protein ThiH	COG1060 Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes
IYC_12669	13,318	13,920	603	hypothetical protein	COG0352 Thiamine monophosphate synthase
IYC_12674	14,255	14,620	366	ArsR family transcriptional regulator	COG0640 Predicted transcriptional regulators
IYC_12679	14,638	15,411	774	nitroreductase	COG1145 Ferredoxin
IYC_12684	15,640	16,119	480	hypothetical protein	COG1396 Predicted transcriptional regulators
IYC_12699	17,089	20,736	3,648	excinuclease ABC subunit A	COG1720 Uncharacterized conserved protein
IYC_12714	22,187	22,933	747	prolyl aminopeptidase	COG0596 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
IYC_12719	23,364	23,651	288	YCII-like protein	COG2350 Uncharacterized protein conserved in bacteria
IYC_12724	23,689	23,823	135	hypothetical protein	
IYC_12729	23,869	24,336	468	glyoxalase/bleomycin resistance protein/dioxygenase	COG0346 Lactoylglutathione lyase and related lyases
IYC_12734	24,594	25,757	1,164	hypothetical protein	COG0789 Predicted transcriptional regulators
IYC_12739	26,297	28,288	1,992	ABC transporter, permease protein	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_12744	28,278	29,045	768	ABC transporter ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_12749	29,405	29,749	345	hypothetical protein	
IYC_12754	29,837	31,531	1,695	hypothetical protein	
IYC_12759	31,909	32,262	354	hypothetical protein	
IYC_12764	32,600	34,693	2,094	PRD domain-containing protein	COG3711 Transcriptional antiterminator
IYC_12769	34,915	36,279	1,365	PTS system ascorbate-specific transporter subunit IIC	COG3037 Uncharacterized protein conserved in bacteria
IYC_12774	36,292	36,576	285	PTS system, IIB component	COG3414 Phosphotransferase system, galactitol-specific IIB component
IYC_12779	36,605	37,042	438	putative PTS system L-ascorbate family, IIA component	COG1762 Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)
IYC_12784	37,066	37,782	717	putative translaldolase	COG0176 Transaldolase
IYC_12789	38,017	38,421	405	hypothetical protein	
IYC_12794	38,759	39,502	744	hypothetical protein	
IYC_12799	39,685	39,858	174	hypothetical protein	
IYC_12804	40,034	40,627	594	membrane protein	COG0344 Predicted membrane protein
IYC_12809	41,346	41,834	489	putative lipoprotein	
IYC_12814	42,193	42,687	495	hypothetical protein	
IYC_12819	42,730	43,344	615	O-acetyltransferase family protein	COG0110 Acetyltransferase (isoleucine patch superfamily)
IYC_12824	43,505	43,597	93	hypothetical protein	
IYC_12829	43,655	44,008	354	M24 family metallopeptidase	COG0006 Xaa-Pro aminopeptidase
IYC_12834	44,092	44,307	216	M24 family metallopeptidase	COG0006 Xaa-Pro aminopeptidase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_12839	44,435	45,841	1,407	amino acid/peptide transporter	COG3104 Dipeptide/tripeptide permease
IYC_12844	45,944	46,684	741	hypothetical protein	COG0500 SAM-dependent methyltransferases
IYC_12849	46,691	47,581	891	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_12854	47,889	48,410	522	hypothetical protein	
IYC_12859	48,521	50,518	1,998	2-enoate reductase	COG1902 NADH:flavin oxidoreductases, Old Yellow Enzyme family
IYC_12864	50,615	51,412	798	MerR family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_12869	51,460	51,753	294	hypothetical protein	
IYC_12874	51,928	52,377	450	HTH-type transcriptional regulator LrpC	COG1522 Transcriptional regulators
IYC_12879	52,516	52,932	417	hypothetical protein	COG2128 Uncharacterized conserved protein
IYC_12884	53,244	54,164	921	PhzF family phenazine biosynthesis protein	COG0384 Predicted epimerase, PhzC/PhzF homolog
IYC_12889	54,488	55,981	1,494	GntR family transcriptional regulator/class I/IIaminotransferase	COG1167 Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs
IYC_12894	56,654	57,109	456	hypothetical protein	
IYC_12899	57,303	58,334	1,032	putative lipase/esterase	COG0657 Esterase/lipase
IYC_12904	58,585	59,013	429	hypothetical protein	
IYC_12909	59,140	60,492	1,353	amidohydrolase domain-containing protein	COG0402 Cytosine deaminase and related metal-dependent hydrolases
IYC_12914	60,699	62,042	1,344	xanthine/uracil permease family protein	COG2233 Xanthine/uracil permeases
IYC_12919	62,612	63,379	768	hypothetical protein	
IYC_12924	63,463	63,639	177	hypothetical protein	
IYC_12929	63,852	64,034	183	hypothetical protein	COG0500 SAM-dependent methyltransferases
IYC_12934	64,161	65,900	1,740	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_12939	66,487	67,806	1,320	amino acid permease family protein	COG0531 Amino acid transporters
IYC_12944	68,071	68,844	774	2-hydroxyglutaryl-CoA dehydratase subunit A	COG1924 Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)
IYC_12949	68,921	70,576	1,656	hypothetical protein	COG3949 Uncharacterized membrane protein
IYC_12954	71,117	72,166	1,050	electron transfer flavoprotein, alpha subunit/FixB family protein	COG2025 Electron transfer flavoprotein, alpha subunit
IYC_12959	72,185	72,970	786	electron transfer flavoprotein subunit beta/FixAfamily protein	COG2086 Electron transfer flavoprotein, beta subunit
IYC_12964	73,010	74,143	1,134	acyl-CoA dehydrogenase family protein	COG1960 Acyl-CoA dehydrogenases
IYC_12969	74,505	75,632	1,128	putative R-2-hydroxyglutaryl-CoA dehydratase, beta subunit	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB
IYC_12974	75,625	76,860	1,236	putative R-2-hydroxyglutaryl-CoA dehydratase, alpha subunit	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB
IYC_12979	76,923	78,137	1,215	putative hydroxyacyl-CoA:R-2-hydroxyglutaryl-CoAtransferase	COG1804 Predicted acyl-CoA transferases/carnitine dehydratase
IYC_12984	78,449	78,601	153	hypothetical protein	
IYC_12989	78,977	79,798	822	iron-sulfur cluster-binding protein	
IYC_12994	79,939	80,094	156	hypothetical protein	
IYC_12999	80,353	81,624	1,272	2-hydroxyglutaryl-CoA dehydratase subunit D	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB
IYC_13004	81,627	83,321	1,695	benzoyl-CoA reductase subunit A	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB
IYC_13009	83,591	84,742	1,152	putative methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_13014	85,405	86,892	1,488	integral membrane sensor signal transduction histidine kinase	COG0642 Signal transduction histidine kinase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_13019	86,885	87,577	693	response regulator receiver	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_13024	87,842	88,300	459	ubiquitin-associated protein	COG1308 Transcription factor homologous to NACalpha-BTF3
IYC_13029	88,539	89,138	600	hypothetical protein	COG3560 Predicted oxidoreductase related to nitroreductase
IYC_13034	89,589	89,849	261	linear amide C-N hydrolase	COG3049 Penicillin V acylase and related amidases
IYC_13039	90,120	90,713	594	GNAT family acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_13044	91,095	91,739	645	hypothetical protein	COG0655 Multimeric flavodoxin WrbA
IYC_13049	91,736	92,242	507	hypothetical protein	COG0655 Multimeric flavodoxin WrbA
IYC_13054	92,403	92,921	519	PadR family transcriptional regulator	COG1695 Predicted transcriptional regulators
IYC_13059	93,043	94,431	1,389	gluconate permease	COG2610 H+/gluconate symporter and related permeases
IYC_13064	94,643	95,647	1,005	4-hydroxythreonine-4-phosphate dehydrogenase	COG1995 Pyridoxal phosphate biosynthesis protein
IYC_13069	95,660	96,970	1,311	hypothetical protein	COG3395 Uncharacterized protein conserved in bacteria
IYC_13074	97,029	97,787	759	DeoR family transcriptional regulator	COG1349 Transcriptional regulators of sugar metabolism
IYC_13079	98,480	99,373	894	CAAX amino terminal protease family protein	COG1266 Predicted metal-dependent membrane protease
IYC_13084	99,785	101,170	1,386	ErfK/YbiS/YcfS/YnhG family protein	COG1376 Uncharacterized protein conserved in bacteria
IYC_13089	101,604	103,334	1,731	peptidoglycan binding protein	COG4932 Predicted outer membrane protein
IYC_13094	104,406	105,077	672	peptidase, sortase like protein	COG3764 Sortase (surface protein transpeptidase)
Genomic Scaffold: NZ_JH470529					
IYC_13099	<1	105	>105	hypothetical protein	
IYC_13104	133	378	246	hypothetical protein	
IYC_13109	339	563	225	hypothetical protein	
IYC_13114	632	>734	>103	hypothetical protein	
Genomic Scaffold: NZ_JH470530					
IYC_13119	624	827	204	D-alanine--poly(phosphoribitol) ligase subunit 1	COG1020 Non-ribosomal peptide synthetase modules and related proteins
IYC_13124	853	963	111	hypothetical protein	
Genomic Scaffold: NZ_JH470531					
IYC_13129	43	1,149	1,107	group 1 glycosyl transferase	COG0438 Glycosyltransferase
IYC_13134	1,169	1,915	747	putative acetyltransferase	COG0110 Acetyltransferase (isoleucine patch superfamily)
IYC_13139	2,014	3,123	1,110	NAD-binding Rossmann fold family oxidoreductase	COG0673 Predicted dehydrogenases and related proteins
IYC_13144	3,186	4,502	1,317	UDP-glucose/GDP-mannose dehydrogenase family protein	COG0677 UDP-N-acetyl-D-mannosaminuronate dehydrogenase
IYC_13149	4,633	6,498	1,866	exopolysaccharide biosynthesis protein	COG1086 Predicted nucleoside-diphosphate sugar epimerases
IYC_13154	6,688	>7946	>1259	putative N-acetylmuramoyl-L-alanine amidase	COG0860 N-acetylmuramoyl-L-alanine amidase
Genomic Scaffold: NZ_JH470532					
IYC_13159	87	620	534	hypothetical protein	
IYC_13164	668	817	150	hypothetical protein	
IYC_13169	878	>1247	>370	phage protein	COG5283 Phage-related tail protein
IYC_13174	<1592	3,611	>2020	phage protein	COG5283 Phage-related tail protein
IYC_13179	3,598	>3954	>357	putative phage tail protein	COG4722 Phage-related protein
Genomic Scaffold: NZ_JH470533					
IYC_13184	67	>481	>415	putative transposase	COG2801 Transposase and inactivated derivatives
Genomic Scaffold: NZ_JH470534					
IYC_13189	174	665	492	hypothetical protein	
IYC_13194	682	927	246	hypothetical protein	COG3655 Predicted transcriptional regulator
IYC_13199	920	2,002	1,083	putative oxidoreductase	COG0644 Dehydrogenases (flavoproteins)
IYC_13204	2,644	2,760	117	hypothetical protein	
IYC_13209	2,840	4,288	1,449	beta-lactamase	COG1680 Beta-lactamase class C and other penicillinbinding proteins
IYC_13214	4,514	5,119	606	DNA-binding protein	COG1396 Predicted transcriptional regulators

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_13219	5,320	5,706	387	hypothetical protein	
IYC_13224	6,139	6,261	123	cell adhesion domain-containing protein	
IYC_13229	6,310	6,522	213	hypothetical protein	COG5492 Bacterial surface proteins containing Ig-like domains
IYC_13234	7,316	8,098	783	hypothetical protein	COG3878 Uncharacterized protein conserved in bacteria
IYC_13239	8,174	8,902	729	hypothetical protein	COG1158 Transcription termination factor
IYC_13244	9,244	10,143	900	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_13249	10,306	10,995	690	GntR family transcriptional regulator	COG2186 Transcriptional regulators
IYC_13254	11,035	12,279	1,245	major facilitator transporter	COG2807 Cyanate permease
IYC_13259	12,388	13,689	1,302	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_13264	14,212	15,354	1,143	iron-containing alcohol dehydrogenase	COG1454 Alcohol dehydrogenase, class IV
IYC_13269	15,457	16,287	831	MerR family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_13274	16,772	17,884	1,113	hypothetical protein	COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily
IYC_13279	17,871	18,137	267	hypothetical protein	
IYC_13284	18,357	18,449	93	hypothetical protein	
IYC_13289	18,452	18,748	297	microcompartments family protein	COG4577 Carbon dioxide concentrating mechanism/carboxysome shell protein
IYC_13294	18,772	19,080	309	microcompartments family protein	COG4577 Carbon dioxide concentrating mechanism/carboxysome shell protein
IYC_13299	19,118	20,611	1,494	aldehyde dehydrogenase family protein	COG1012 NAD-dependent aldehyde dehydrogenases
IYC_13304	20,661	23,201	2,541	formate C-acetyltransferase	COG1882 Pyruvate-formate lyase
IYC_13309	23,365	24,318	954	glycyl-radical enzyme activating protein family	COG1180 Pyruvate-formate lyase-activating enzyme
IYC_13314	24,429	24,782	354	ethanolamine utilization protein EutS-like protein	COG4810 Ethanolamine utilization protein
IYC_13319	24,779	25,225	447	GTP-binding protein, EutP/PduV family	COG4917 Ethanolamine utilization protein
IYC_13324	25,232	25,960	729	hypothetical protein	
IYC_13329	25,942	26,787	846	ethanolamine utilization protein EutJ	COG4820 Ethanolamine utilization protein, possible chaperonin
IYC_13334	26,865	27,167	303	hypothetical protein	
IYC_13339	27,164	27,427	264	ethanolamine utilization protein EutN/carboxysome structural protein Ccml	COG4576 Carbon dioxide concentrating mechanism/carboxysome shell protein
IYC_13344	27,626	28,288	663	ethanolamine utilization protein EutQ-like protein	COG4766 Ethanolamine utilization protein
IYC_13349	28,309	29,112	804	microcompartments family protein	COG4577 Carbon dioxide concentrating mechanism/carboxysome shell protein
IYC_13354	29,147	30,745	1,599	acetaldehyde dehydrogenase	COG1012 NAD-dependent aldehyde dehydrogenases
IYC_13359	30,745	31,395	651	propanediol utilization phosphotransacylase	COG4869 Propanediol utilization protein
IYC_13364	31,469	31,759	291	microcompartments family protein	COG4577 Carbon dioxide concentrating mechanism/carboxysome shell protein
IYC_13369	32,198	32,407	210	hypothetical protein	
IYC_13374	32,419	>32550	>132	hypothetical protein	
Genomic Scaffold: NZ_JH470535					
IYC_13379	177	917	741	hypothetical protein	COG0217 Uncharacterized conserved protein
IYC_13384	1,200	1,847	648	hypothetical protein	COG1739 Uncharacterized conserved protein
IYC_13389	1,863	2,525	663	PBS lyase HEAT-like repeat-containing protein	COG1413 FOG: HEAT repeat
IYC_13394	2,617	4,035	1,419	GntR family transcriptional regulator	COG1167 Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs
IYC_13399	4,289	5,182	894	nucleotidyltransferase domain-containing protein	COG1708 Predicted nucleotidyltransferases
IYC_13404	5,332	5,553	222	hypothetical protein	
IYC_13409	5,728	5,886	159	rubredoxin	COG1773 Rubredoxin
IYC_13414	6,015	7,799	1,785	GTP-binding protein	COG2262 GTPases

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_13419	7,941	8,468	528	hypoxanthine phosphoribosyltransferase	COG0634 Hypoxanthine-guanine phosphoribosyltransferase
IYC_13424	8,673	9,227	555	sporulation protein YunB	
IYC_13429	9,499	11,994	2,496	1A family penicillin-binding protein	COG0744 Membrane carboxypeptidase (penicillin-binding protein)
IYC_13434	12,023	12,379	357	SpoVA family protein	
IYC_13439	12,394	13,401	1,008	stage V sporulation protein AD	
IYC_13444	13,424	13,879	456	SpoVA family protein	
IYC_13449	14,082	14,837	756	sporulation sigma factor SigF	COG1191 DNA-directed RNA polymerase specialized sigma subunit
IYC_13454	14,856	15,278	423	anti-sigma F factor	COG2172 Anti-sigma regulatory factor (Ser/Thr protein kinase)
IYC_13459	15,295	15,630	336	anti-sigma F factor antagonist	COG1366 Anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)
IYC_13464	15,884	16,711	828	spermidine synthase	COG0421 Spermidine synthase
IYC_13469	16,939	18,672	1,734	modification methylase family protein	COG1002 Type II restriction enzyme, methylase subunits
IYC_13474	18,739	19,725	987	polysaccharide deacetylase family protein	COG0726 Predicted xylanase/chitin deacetylase
IYC_13479	19,863	20,411	549	nitroreductase	COG0778 Nitroreductase
IYC_13484	20,577	21,293	717	WecB/TagA/CpsF family glycosyl transferase	COG1922 Teichoic acid biosynthesis proteins
IYC_13489	21,491	22,144	654	sugar transferase family protein	COG2148 Sugar transferases involved in lipopolysaccharide synthesis
IYC_13494	22,164	23,588	1,425	putative polysaccharide transporter	COG2244 Membrane protein involved in the export of O-antigen and teichoic acid
IYC_13499	23,766	25,322	1,557	integral membrane protein MviN	COG0728 Uncharacterized membrane protein, putative virulence factor
IYC_13504	25,341	26,447	1,107	glycosyl transferase, group 1 family protein	COG0438 Glycosyltransferase
IYC_13509	26,644	27,891	1,248	exopolysaccharide biosynthesis family protein	COG3307 Lipid A core - O-antigen ligase and relatedenzymes
IYC_13514	27,973	29,898	1,926	hypothetical protein	COG2247 Putative cell wall-binding domain
IYC_13519	29,923	31,116	1,194	group 1 glycosyl transferase	COG0438 Glycosyltransferase
IYC_13524	31,253	32,287	1,035	succinoglycan biosynthesis protein exoA	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_13529	32,305	33,732	1,428	hypothetical protein	
IYC_13534	33,755	35,092	1,338	sugar-binding protein	COG1653 ABC-type sugar transport system, periplasmic component
IYC_13539	35,113	>35479	>367	UDP-N-acetylglucosamine 2-epimerase	COG0381 UDP-N-acetylglucosamine 2-epimerase
Genomic Scaffold: NZ_JH470536					
IYC_13544	1,428	1,688	261	hypothetical protein	
IYC_13549	1,702	1,851	150	hypothetical protein	
IYC_13554	1,944	2,159	216	hypothetical protein	
IYC_13559	2,161	2,481	321	hypothetical protein	
IYC_13564	2,497	3,024	528	hypothetical protein	
IYC_13569	3,035	3,508	474	hypothetical protein	
IYC_13574	3,505	4,650	1,146	hypothetical protein	COG3096 Uncharacterized protein involved in chromosome partitioning
IYC_13579	4,650	5,744	1,095	phage-like protein	
IYC_13584	5,747	6,637	891	phage-like protein	
IYC_13589	6,624	12,353	5,730	phage tail tape measure protein, family, core region	COG5283 Phage-related tail protein
IYC_13594	12,553	13,173	621	hypothetical protein	
IYC_13599	13,237	13,413	177	hypothetical protein	
IYC_13604	13,403	13,822	420	hypothetical protein	
IYC_13609	13,878	14,504	627	hypothetical protein	
IYC_13614	14,504	14,842	339	hypothetical protein	
IYC_13619	14,842	15,216	375	hypothetical protein	
IYC_13624	15,209	15,574	366	hypothetical protein	
IYC_13629	15,571	15,852	282	hypothetical protein	
IYC_13634	15,855	16,058	204	hypothetical protein	
IYC_13639	16,114	17,148	1,035	phage protein	
IYC_13644	17,169	17,525	357	hypothetical protein	
IYC_13649	17,539	18,177	639	putative phage scaffold protein	
IYC_13654	18,228	18,446	219	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_13659	18,436	18,732	297	hypothetical protein	
IYC_13664	18,743	19,702	960	hypothetical protein	
IYC_13669	19,702	21,174	1,473	hypothetical protein	
IYC_13674	21,174	22,631	1,458	hypothetical protein	
IYC_13679	22,624	23,307	684	phage protein	COG5484 Uncharacterized conserved protein
IYC_13684	23,374	23,796	423	hypothetical protein	
Genomic Scaffold: NZ_JH470537					
IYC_13689	<1	499	>499	Collagen triple helix repeat protein	
IYC_13694	885	>1354	>470	Collagen triple helix repeat-containing protein	
IYC_13699	1,814	2,890	1,077	beta 1,4 glucosyltransferase	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_13704	3,223	3,810	588	hypothetical protein	COG1217 Predicted membrane GTPase involved in stress response
IYC_13709	3,958	5,721	1,764	hypothetical protein	
IYC_13714	5,926	6,825	900	hypothetical protein	
IYC_13719	7,318	7,431	114	hypothetical protein	COG1088 dTDP-D-glucose 4,6-dehydratase
IYC_13724	7,646	10,012	2,367	U32 family peptidase	COG0826 Collagenase and related proteases
IYC_13729	10,329	12,695	2,367	recombination and DNA strand exchange inhibitor protein	COG1193 Mismatch repair ATPase (MutS family)
IYC_13734	12,880	13,338	459	hypothetical protein	COG1683 Uncharacterized conserved protein
IYC_13739	13,938	15,770	1,833	phosphoglycerol transferase	COG1368 Phosphoglycerol transferase and related proteins, alkaline phosphatase superfamily
IYC_13744	15,773	16,009	237	hypothetical protein	
IYC_13749	16,209	17,615	1,407	glutamate decarboxylase	COG0076 Glutamate decarboxylase and related PLP-dependent proteins
IYC_13754	17,929	18,942	1,014	dehydrogenase	COG1304 L-lactate dehydrogenase (FMN-dependent) andrelated alpha-hydroxy acid dehydrogenases
IYC_13759	19,402	20,304	903	phosphate acetyl/butyryl transferase	COG0280 Phosphotransacetylase
IYC_13764	20,679	21,518	840	hypothetical protein	
IYC_13769	21,676	23,136	1,461	Orn/Lys/Arg decarboxylase	COG1982 Arginine/lysine/ornithine decarboxylases
IYC_13774	23,339	23,875	537	50S ribosomal protein L25/general stress proteinCtc	COG1825 Ribosomal protein L25 (general stress protein Ctc)
IYC_13779	24,570	26,297	1,728	phosphoglucomutase/phosphomannomutase family protein	COG1109 Phosphomannomutase
IYC_13784	26,442	27,518	1,077	TPR domain-containing protein	COG0457 FOG: TPR repeat
IYC_13789	27,568	28,464	897	UTP-glucose-1-phosphate uridylyltransferase	COG1210 UDP-glucose pyrophosphorylase
IYC_13794	28,581	>28681	>101	hypothetical protein	
Genomic Scaffold: NZ_JH470538					
IYC_13799	<1	412	>412	hypothetical protein	
IYC_13804	<388	719	>332	hypothetical protein	COG1804 Predicted acyl-CoA transferases/carnitine dehydratase
IYC_13809	719	3,592	2,874	hypothetical protein	
IYC_13814	<3534	3,686	>153	hypothetical protein	
IYC_13819	3,686	3,832	147	hypothetical protein	
Genomic Scaffold: NZ_JH470539					
IYC_13824	<1	207	>207	hypothetical protein	COG1804 Predicted acyl-CoA transferases/carnitine dehydratase
Genomic Scaffold: NZ_JH470541					
IYC_13833	73	>203	>131	hypothetical protein	
Genomic Scaffold: NZ_JH470542					
IYC_13838	<1	>231	>231	spore germination protein	
Genomic Scaffold: NZ_JH470543					
IYC_13843	129	1,190	1,062	tRNA-specific 2-thiouridylase MnmA	COG0482 Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase, contains the PP-loop ATPase domain
IYC_13848	1,511	3,811	2,301	putative helicase	COG3973 Superfamily I DNA and RNA helicases
IYC_13853	4,611	>5391	>781	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter

locus_tag	Minimum	Maximum	Length	Product	COG Group
Genomic Scaffold: NZ_JH470544					
IYC_13858	<1	>1203	>1203	phage endopeptidase	COG4926 Phage-related protein
Genomic Scaffold: NZ_JH470545					
IYC_13863	262	513	252	putative lipoprotein	
IYC_13868	524	709	186	phage protein	
IYC_13873	882	1,004	123	hypothetical protein	
IYC_13878	1,142	1,987	846	N-acetylmuramoyl-L-alanine amidase	COG0860 N-acetylmuramoyl-L-alanine amidase
IYC_13883	2,029	>2213	>185	hypothetical protein	
Genomic Scaffold: NZ_JH470546					
IYC_13888	<1	205	>205	putative N-acetylmuramoyl-L-alanine amidase	overlaps another CDS with the same product name
IYC_13893	162	875	714	putative N-acetylmuramoyl-L-alanine amidase	COG0860 N-acetylmuramoyl-L-alanine amidase; overlaps another CDS with the same product name
IYC_13898	1,049	>1778	>730	putative N-acetylmuramoyl-L-alanine amidase	
Genomic Scaffold: NZ_JH470547					
IYC_13903	18	>740	>723	UDP-N-acetylglucosamine 2-epimerase	COG0381 UDP-N-acetylglucosamine 2-epimerase
Genomic Scaffold: NZ_JH470548					
IYC_13908	40	1,521	1,482	stage IV sporulation protein A	
IYC_13913	1,999	2,469	471	hypothetical protein	
IYC_13918	2,540	3,538	999	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase	COG0240 Glycerol-3-phosphate dehydrogenase
IYC_13923	3,705	5,024	1,320	GTP-binding protein Der	COG1160 Predicted GTPases
IYC_13928	5,027	6,361	1,335	hypothetical protein	COG1625 Fe-S oxidoreductase, related to NifB/MoaA family
IYC_13933	6,780	7,433	654	phosphate transport system regulatory protein PhoU	COG0704 Phosphate uptake regulator
IYC_13938	7,463	8,212	750	phosphate transporter ATP-binding protein	COG1117 ABC-type phosphate transport system, ATPasecomponent
IYC_13943	8,225	9,109	885	phosphate ABC transporter, permease protein PstA	COG0581 ABC-type phosphate transport system, permease component
IYC_13948	9,109	10,032	924	phosphate ABC transporter, permease protein PstC	COG0573 ABC-type phosphate transport system, permease component
IYC_13953	10,238	11,110	873	phosphate ABC transporter, phosphate-binding protein	COG0226 ABC-type phosphate transport system, periplasmic component
IYC_13958	11,357	13,066	1,710	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_13963	13,078	13,776	699	putative phosphate regulon DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_13968	13,787	14,506	720	multicopper polyphenol oxidase	COG1496 Uncharacterized conserved protein
IYC_13973	14,782	15,237	456	transcriptional regulator NrdR	COG1327 Predicted transcriptional regulator, consists of a Zn-ribbon and ATP-cone domains
IYC_13978	15,369	15,632	264	PRC-barrel domain-containing protein	COG1873 Uncharacterized conserved protein
IYC_13983	15,778	16,551	774	sporulation sigma factor SigG	COG1191 DNA-directed RNA polymerase specialized sigma subunit
IYC_13988	16,626	17,333	708	sporulation sigma factor SigE	COG1191 DNA-directed RNA polymerase specialized sigma subunit
IYC_13993	17,345	18,142	798	sporulation factor SpoIIGA	
IYC_13998	18,478	19,587	1,110	cell division protein FtsZ	COG0206 Cell division GTPase
IYC_14003	19,610	20,863	1,254	cell division protein FtsA	COG0849 Actin-like ATPase involved in cell division
IYC_14008	21,019	22,293	1,275	glycosyl transferase, group 2 family protein	COG1215 Glycosyltransferases, probably involved in cell wall biogenesis
IYC_14013	22,732	22,866	135	hypothetical protein	
IYC_14018	22,977	24,029	1,053	twitching motility protein	COG2805 Tfp pilus assembly protein, pilus retraction ATPase PilT
IYC_14023	24,031	24,807	777	shikimate 5-dehydrogenase	COG0169 Shikimate 5-dehydrogenase
IYC_14028	24,946	25,650	705	sporulation sigma factor SigK	COG1191 DNA-directed RNA polymerase specialized sigma subunit
IYC_14033	26,430	26,669	240	hypothetical protein	
IYC_14038	26,681	27,766	1,086	hypothetical protein	COG4127 Uncharacterized conserved protein
IYC_14043	27,888	28,655	768	N-acetylmuramoyl-L-alanine amidase	COG0860 N-acetylmuramoyl-L-alanine amidase
IYC_14048	28,703	28,897	195	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_14053	28,901	29,179	279	hypothetical protein	
IYC_14058	29,266	29,409	144	hypothetical protein	
IYC_14063	29,409	29,609	201	hypothetical protein	
IYC_14068	29,622	31,184	1,563	hypothetical protein	
IYC_14073	32,201	32,317	117	hypothetical protein	
IYC_14078	32,339	>32698	>360	hypothetical protein	COG1804 Predicted acyl-CoA transferases/carnitine dehydratase
Genomic Scaffold: NZ_JH470549					
IYC_14083	422	2,203	1,782	single-stranded-DNA-specific exonuclease RecJ	COG0608 Single-stranded DNA-specific exonuclease
IYC_14088	2,382	3,200	819	AP endonuclease	COG0648 Endonuclease IV
IYC_14093	3,490	4,893	1,404	NHL repeat-containing protein	COG0526 Thiol-disulfide isomerase and thioredoxins
IYC_14098	5,344	6,615	1,272	putative UV damage endonuclease	COG4294 UV damage repair endonuclease
IYC_14103	6,758	7,975	1,218	M16 family peptidase	COG0612 Predicted Zn-dependent peptidases
IYC_14108	8,255	8,665	411	hypothetical protein	COG2246 Predicted membrane protein
IYC_14113	8,733	9,284	552	DJ-1 family protein	COG0693 Putative intracellular protease/amidase
IYC_14118	9,570	10,133	564	glycerol uptake operon antiterminator regulatoryprotein	COG1954 Glycerol-3-phosphate responsive antiterminator (mRNA-binding)
IYC_14123	10,182	13,052	2,871	DNA topoisomerase IV subunit A	COG0188 Type IIA topoisomerase (DNA gyrase/topo II,topoisomerase IV), A subunit
IYC_14128	13,128	15,083	1,956	DNA topoisomerase IV subunit B	COG0187 Type IIA topoisomerase (DNA gyrase/topo II,topoisomerase IV), B subunit
IYC_14133	15,696	16,073	378	hypothetical protein	
IYC_14138	16,266	17,438	1,173	transporter, monovalent cation:proton antiporter-2 family protein	COG0475 Kef-type K+ transport systems, membrane components
IYC_14143	17,451	17,801	351	hypothetical protein	
IYC_14148	18,064	18,300	237	hypothetical protein	
IYC_14153	18,320	19,213	894	polysaccharide deacetylase family protein	COG0726 Predicted xylanase/chitin deacetylase
IYC_14158	19,235	20,608	1,374	Na+ driven multidrug efflux pump	COG0534 Na+-driven multidrug efflux pump
IYC_14163	20,901	22,190	1,290	hypothetical protein	COG0823 Periplasmic component of the Tol biopolymertransport system
IYC_14168	22,228	23,208	981	UDP-glucose 4-epimerase	COG1087 UDP-glucose 4-epimerase
IYC_14173	23,564	23,980	417	flavodoxin	COG0716 Flavodoxins
IYC_14178	24,247	25,668	1,422	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_14183	26,308	28,374	2,067	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_14188	28,782	30,278	1,497	glycerol kinase	COG0554 Glycerol kinase
IYC_14193	30,569	31,273	705	glycerol uptake facilitator protein	COG0580 Glycerol uptake facilitator and related permeases (Major Intrinsic Protein Family)
IYC_14198	31,815	32,135	321	hypothetical protein	COG0393 Uncharacterized conserved protein
IYC_14203	32,249	32,935	687	VanZ family membrane protein	COG4767 Glycopeptide antibiotics resistance protein
IYC_14208	33,080	34,156	1,077	ABC transporter permease	COG1277 ABC-type transport system involved in multi-copper enzyme maturation, permease component
IYC_14213	34,140	35,036	897	ABC transporter ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_14218	35,042	35,422	381	GntR family transcriptional regulator	COG1725 Predicted transcriptional regulators
IYC_14223	35,589	36,104	516	VanZ family membrane protein	COG4767 Glycopeptide antibiotics resistance protein
IYC_14228	36,535	38,247	1,713	hybrid cluster protein	COG1151 6Fe-6S prismane cluster-containing protein
IYC_14233	38,534	40,252	1,719	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_14238	40,511	41,101	591	membrane-spanning protein	
IYC_14243	41,525	42,757	1,233	aminopeptidase II	COG2309 Leucyl aminopeptidase (aminopeptidase T)
IYC_14248	43,030	44,175	1,146	spore germination protein, GerAC	
IYC_14253	44,239	45,264	1,026	spore germination protein	COG0531 Amino acid transporters

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_14258	45,254	46,741	1,488	spore germination protein, GerAA	
IYC_14263	47,010	47,219	210	hypothetical protein	
IYC_14268	47,488	48,069	582	glutamine amidotransferase subunit PdxT	COG0311 Predicted glutamine amidotransferase involved in pyridoxine biosynthesis
IYC_14273	48,071	48,943	873	pyridoxal biosynthesis lyase PdxS	COG0214 Pyridoxine biosynthesis enzyme
IYC_14278	49,310	50,041	732	polysaccharide deacetylase family protein	COG0726 Predicted xylanase/chitin deacetylase
IYC_14283	50,232	51,809	1,578	DEAD-box ATP dependent DNA helicase	COG0513 Superfamily II DNA and RNA helicases
IYC_14288	51,886	52,053	168	hypothetical protein	
IYC_14293	52,502	52,972	471	hypothetical protein	
IYC_14298	53,099	53,509	411	universal stress protein family protein	COG0589 Universal stress protein UspA and related nucleotide-binding proteins
IYC_14303	54,270	54,911	642	iron-sulfur cluster-binding protein	COG0348 Polyferredoxin
IYC_14308	55,025	55,501	477	hypothetical protein	COG0081 Ribosomal protein L1
IYC_14313	55,607	56,524	918	glutaminase	COG2066 Glutaminase
IYC_14318	56,896	57,927	1,032	hypothetical protein	COG3883 Uncharacterized protein conserved in bacteria
IYC_14323	58,461	58,598	138	putative lipoprotein	
IYC_14328	58,765	60,825	2,061	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_14333	61,357	62,157	801	hypothetical protein	
IYC_14338	62,519	63,547	1,029	ribonucleoside-diphosphate reductase, beta subunit	COG0208 Ribonucleotide reductase, beta subunit
IYC_14343	63,741	66,032	2,292	ribonucleotide-diphosphate reductase subunit alpha	COG0209 Ribonucleotide reductase, alpha subunit
IYC_14348	66,530	67,705	1,176	GAF sensor signal transduction histidine kinase	COG0642 Signal transduction histidine kinase
IYC_14353	67,983	69,113	1,131	hypothetical protein	
IYC_14358	69,283	70,071	789	hypothetical protein	COG3968 Uncharacterized protein related to glutamine synthetase
IYC_14363	70,271	71,320	1,050	DNA polymerase IV	COG0389 Nucleotidyltransferase/DNA polymerase involved in DNA repair
IYC_14368	71,923	72,327	405	putative chloride channel protein EriC	
IYC_14373	72,403	73,638	1,236	dihydropyrimidine dehydrogenase subunit B	COG0167 Dihydroorotate dehydrogenase
IYC_14378	73,638	74,954	1,317	pyridine nucleotide-disulfide oxidoreductase family protein	COG0493 NADPH-dependent glutamate synthase beta chain and related oxidoreductases
IYC_14383	75,410	75,958	549	pyrimidine operon regulatory protein/uracil phosphoribosyltransferase	COG2065 Pyrimidine operon attenuation protein/uracil phosphoribosyltransferase
IYC_14388	76,280	77,341	1,062	hypothetical protein	COG1434 Uncharacterized conserved protein
IYC_14393	77,936	78,355	420	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_14398	78,373	80,631	2,259	putative excinuclease ABC, A subunit	COG0178 Excinuclease ATPase subunit
IYC_14403	80,923	81,444	522	2'-5' RNA ligase	COG1514 2'-5' RNA ligase
IYC_14408	81,563	81,961	399	hypothetical protein	
IYC_14413	82,471	83,922	1,452	hypothetical protein	
IYC_14418	84,279	84,446	168	putative flagellin B	
IYC_14423	84,687	85,124	438	MutT/nudix family protein	COG0494 NTP pyrophosphohydrolases including oxidative damage repair enzymes
IYC_14428	85,661	86,128	468	FUR family transcriptional regulator	COG0735 Fe2+/Zn2+ uptake regulation proteins
IYC_14433	86,261	88,270	2,010	hypothetical protein	
IYC_14438	88,366	90,585	2,220	glucosyl hydrolase family protein	COG3325 Chitinase
IYC_14443	90,812	91,546	735	glucosamine-6-phosphate deaminase	COG0363 6-phosphogluconolactonase/Glucosamine-6-phosphate isomerase/deaminase
IYC_14448	91,629	92,771	1,143	N-acetylglucosamine-6-phosphate deacetylase	COG1820 N-acetylglucosamine-6-phosphate deacetylase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_14453	92,773	93,666	894	N-acetylmuramic acid-6-phosphate etherase	COG2103 Predicted sugar phosphate isomerase
IYC_14458	93,723	94,457	735	GntR family transcriptional regulator	COG2188 Transcriptional regulators
IYC_14463	94,790	95,647	858	BglG family transcriptional antiterminator	COG3711 Transcriptional antiterminator
IYC_14468	95,863	96,342	480	PTS system, glucose subfamily, IIA component	COG2190 Phosphotransferase system IIA components
IYC_14473	97,034	98,476	1,443	PTS system, N-acetylglucosamine-specific, IIBC component	COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific
IYC_14478	98,725	98,922	198	hypothetical protein	COG1476 Predicted transcriptional regulators
IYC_14483	99,210	100,580	1,371	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_14488	100,741	102,135	1,395	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_14493	102,371	103,570	1,200	metallo-beta-lactamase family protein/flavodoxin	COG0426 Uncharacterized flavoproteins
IYC_14498	103,628	105,535	1,908	acyl-CoA dehydrogenase family protein/electron transfer protein	COG1960 Acyl-CoA dehydrogenases
IYC_14503	105,726	106,730	1,005	hypothetical protein	
IYC_14508	106,767	107,108	342	hydrogenase nickel insertion protein HypA	COG0375 Zn finger protein HypA/HybF (possibly regulating hydrogenase expression)
IYC_14513	107,164	109,842	2,679	pyridine nucleotide-disulfide oxidoreductase family protein	COG1145 Ferredoxin
IYC_14518	110,089	110,760	672	hydrogenase accessory protein HypB	COG0378 Ni2+-binding GTPase involved in regulation of expression and maturation of urease and hydrogenase
IYC_14523	111,018	111,626	609	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_14528	112,267	114,846	2,580	ABC transporter ATP-binding protein	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_14533	114,843	115,526	684	lipoprotein-releasing system ATP-binding proteinLolD	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_14538	115,610	116,599	990	sensor protein ResE	COG0642 Signal transduction histidine kinase
IYC_14543	116,586	117,272	687	hypothetical protein	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_14548	117,269	117,463	195	putative transposase	
IYC_14553	117,829	119,184	1,356	23S rRNA (uracil-5-)-methyltransferase RumA	COG2265 SAM-dependent methyltransferases related totRNA (uracil-5-)-methyltransferase
IYC_14558	119,412	119,870	459	hypothetical protein	
IYC_14563	120,205	122,064	1,860	sulfatase family protein	COG1368 Phosphoglycerol transferase and related proteins, alkaline phosphatase superfamily
IYC_14568	122,095	122,847	753	hypothetical protein	COG0730 Predicted permeases
IYC_14573	123,181	123,645	465	6,7-dimethyl-8-ribityllumazine synthase	COG0054 Riboflavin synthase beta-chain
IYC_14578	123,677	124,882	1,206	bifunctional riboflavin biosynthesis protein RibAB	COG0108 3,4-dihydroxy-2-butanone 4-phosphate synthase
IYC_14583	125,000	125,653	654	riboflavin synthase subunit alpha	COG0307 Riboflavin synthase alpha chain
IYC_14588	125,832	126,926	1,095	riboflavin biosynthesis protein RibD	COG0117 Pyrimidine deaminase
IYC_14593	127,328	128,764	1,437	aminoacyl-histidine dipeptidase	COG2195 Di- and tripeptidases
IYC_14598	129,157	130,050	894	metallo-beta-lactamase family protein	COG0491 Zn-dependent hydrolases, including glyoxylases
IYC_14603	130,268	132,913	2,646	magnesium-translocating P-type ATPase	COG0474 Cation transport ATPase
IYC_14608	133,094	133,213	120	hypothetical protein	
IYC_14613	133,506	133,910	405	hypothetical protein	
IYC_14618	134,017	135,099	1,083	spore germination protein	
IYC_14623	135,223	136,629	1,407	GerA family spore germination protein	
IYC_14628	136,720	137,814	1,095	spore germination protein	
IYC_14633	137,970	139,211	1,242	phosphoribosylamine--glycine ligase	COG0151 Phosphoribosylamine-glycine ligase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_14638	139,374	140,873	1,500	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	COG0138 AICAR transformylase/IMP cyclohydrolase PurH (only IMP cyclohydrolase domain in Aful)
IYC_14643	141,093	141,710	618	phosphoribosylglycinamide formyltransferase	COG0299 Folate-dependent phosphoribosylglycinamide formyltransferase PurN
IYC_14648	141,838	142,833	996	phosphoribosylaminoimidazole synthetase	COG0150 Phosphoribosylaminoimidazole (AIR) synthetase
IYC_14653	143,005	144,453	1,449	amidophosphoribosyltransferase	COG0034 Glutamine phosphoribosylpyrophosphate amidotransferase
IYC_14658	144,568	145,272	705	phosphoribosylaminoimidazole-succinocarboxamide synthase	COG0152 Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase
IYC_14663	145,272	145,751	480	phosphoribosylaminoimidazole carboxylase, catalytic subunit	COG0041 Phosphoribosylcarboxyaminoimidazole (NCAIR)mutase
IYC_14668	146,409	148,964	2,556	xanthine dehydrogenase family protein molybdopterin-binding subunit	COG2080 Aerobic-type carbon monoxide dehydrogenase,small subunit CoxS/CutS homologs
IYC_14673	149,314	150,531	1,218	peptidase	COG0624 Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases
IYC_14678	150,557	151,885	1,329	putative chlorohydrolase/aminohydrolase	COG0402 Cytosine deaminase and related metal-dependent hydrolases
IYC_14683	151,916	154,933	3,018	putative selenate reductase subunit YgfK	COG0493 NADPH-dependent glutamate synthase beta chain and related oxidoreductases
IYC_14688	154,970	156,166	1,197	aspartate/ornithine carbamoyltransferase family protein	COG0078 Ornithine carbamoyltransferase
IYC_14693	156,529	157,902	1,374	phenylhydantoinase	COG0044 Dihydroorotase and related cyclic amidohydrolases
IYC_14698	157,985	159,358	1,374	xanthine/uracil permease family protein	COG2233 Xanthine/uracil permeases
IYC_14703	159,381	160,592	1,212	diaminopropionate ammonia-lyase	COG1171 Threonine dehydratase
IYC_14708	160,880	162,643	1,764	sigma-54 dependent transcriptional regulator	COG3829 Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding domains
IYC_14713	163,011	163,748	738	hypothetical protein	
IYC_14718	163,763	164,377	615	hypothetical protein	COG2068 Uncharacterized MobA-related protein
IYC_14723	164,358	165,206	849	hypothetical protein	COG0845 Membrane-fusion protein
IYC_14728	165,188	165,997	810	putative xanthine dehydrogenase accessory factor	COG1975 Xanthine and CO dehydrogenases maturation factor, XdhC/CoxF family
IYC_14733	166,023	166,568	546	hypothetical protein	COG1618 Predicted nucleotide kinase
IYC_14738	166,565	167,533	969	iron chelate uptake ABC transporter permease	COG0609 ABC-type Fe3+-siderophore transport system,permease component
IYC_14743	167,530	168,324	795	iron chelate ABC transporter ATP-binding protein	COG1120 ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components
IYC_14748	168,515	169,573	1,059	putative iron chelate uptake ABC transporter, solute-binding protein	COG0614 ABC-type Fe3+-hydroxamate transport system,periplasmic component
IYC_14753	169,711	170,043	333	hypothetical protein	
IYC_14758	170,173	170,745	573	lipase/acylhydrolase, GDSL family protein	COG2755 Lysophospholipase L1 and related esterases
IYC_14763	170,810	171,355	546	hypothetical protein	COG1434 Uncharacterized conserved protein
IYC_14768	171,563	171,676	114	hypothetical protein	
IYC_14773	171,785	173,227	1,443	xanthine/uracil permease family protein	COG2252 Permeases
IYC_14778	173,759	>173997	>239	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter
Genomic Scaffold: NZ_JH470550					
IYC_14783	<1	235	>235	hypothetical protein	COG4545 Glutaredoxin-related protein
Genomic Scaffold: NZ_JH470551					
IYC_14788	<1	>1203	>1203	phage endopeptidase	COG4926 Phage-related protein
Genomic Scaffold: NZ_JH470552					
IYC_14793	<1	132	>132	hypothetical protein	
IYC_14798	144	254	111	hypothetical protein	
IYC_14803	338	640	303	hypothetical protein	
IYC_14808	649	1,317	669	hypothetical protein	COG4495 Uncharacterized protein conserved in bacteria
IYC_14813	1,524	3,752	2,229	hypothetical protein	COG1511 Predicted membrane protein
IYC_14818	3,766	8,118	4,353	FtsK/SpoIIIE family DNA segregation ATPase	COG1674 DNA segregation ATPase FtsK/SpoIIIE and related proteins
IYC_14823	8,120	9,259	1,140	hypothetical protein	COG4499 Predicted membrane protein
IYC_14828	9,261	9,494	234	hypothetical protein	
IYC_14833	9,466	9,954	489	hypothetical protein	
IYC_14838	10,272	10,391	120	hypothetical protein	
IYC_14843	10,574	10,894	321	hypothetical protein	COG1733 Predicted transcriptional regulators
IYC_14848	11,047	11,679	633	hypothetical protein	COG2910 Putative NADH-flavin reductase
IYC_14853	11,864	12,034	171	hypothetical protein	
IYC_14858	12,725	13,333	609	hypothetical protein	COG0613 Predicted metal-dependent phosphoesterases (PHP family)
IYC_14863	13,633	13,908	276	hypothetical protein	COG2197 Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain
IYC_14868	13,989	15,023	1,035	chloramphenicol/florfenicol resistance protein	COG0820 Predicted Fe-S-cluster redox enzyme
IYC_14873	15,186	16,664	1,479	drug resistance ABC transporter ATP-binding protein	COG0488 ATPase components of ABC transporters with duplicated ATPase domains
IYC_14878	17,128	18,507	1,380	23S rRNA (uracil-5-)-methyltransferase RumA	COG2265 SAM-dependent methyltransferases related totRNA (uracil-5-)-methyltransferase
IYC_14883	18,734	19,090	357	hypothetical protein	
IYC_14888	19,421	20,431	1,011	aldo/keto reductase family oxidoreductase	COG0667 Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)
IYC_14893	20,577	22,334	1,758	pyruvate kinase	COG0469 Pyruvate kinase
IYC_14898	22,368	23,327	960	6-phosphofructokinase	COG0205 6-phosphofructokinase
IYC_14903	23,570	27,127	3,558	DNA polymerase III DnaE	COG0587 DNA polymerase III, alpha subunit
IYC_14908	27,287	28,234	948	putative sporulation transcription regulator whiA	COG1481 Uncharacterized protein conserved in bacteria
IYC_14913	28,241	29,578	1,338	hypothetical protein	COG0391 Uncharacterized conserved protein
IYC_14918	29,575	30,459	885	glmZ(sRNA)-inactivating NTPase	COG1660 Predicted P-loop-containing kinase
IYC_14923	30,579	31,499	921	UDP-N-acetylenolpyruvoylglucosamine reductase	COG0812 UDP-N-acetylmuramate dehydrogenase
IYC_14928	31,568	32,299	732	putative hydrolase	COG1387 Histidinol phosphatase and related hydrolases of the PHP family
IYC_14933	32,359	34,215	1,857	excinuclease ABC subunit C	COG0322 Nuclease subunit of the excinuclease complex
IYC_14938	34,229	34,837	609	putative metal-dependent hydrolase, membrane-bound	COG1988 Predicted membrane-bound metal-dependent hydrolases
IYC_14943	35,028	35,924	897	HD domain-containing protein	COG3481 Predicted HD-superfamily hydrolase
IYC_14948	35,953	37,380	1,428	penicillin-binding protein	COG0768 Cell division protein FtsI/penicillin-binding protein 2
IYC_14953	37,385	38,590	1,206	cell cycle protein FtsW	COG0772 Bacterial cell division membrane protein
IYC_14958	38,625	39,062	438	FHA domain-containing protein	COG1716 FOG: FHA domain
IYC_14963	39,161	41,983	2,823	excinuclease ABC subunit A	COG0178 Excinuclease ATPase subunit
IYC_14968	42,002	43,990	1,989	excinuclease ABC subunit B	COG0556 Helicase subunit of the DNA excision repaircomplex
IYC_14973	43,992	45,287	1,296	hypothetical protein	COG0265 Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain
IYC_14978	45,375	46,580	1,206	carboxyl-terminal protease	COG0793 Periplasmic protease
IYC_14983	46,790	47,680	891	putative cell division protein FtsX	COG2177 Cell division protein
IYC_14988	47,670	48,356	687	putative cell division ATP-binding protein FtsE	COG2884 Predicted ATPase involved in cell division

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_14993	48,512	49,414	903	hypothetical protein	COG1284 Uncharacterized conserved protein
IYC_14998	49,502	49,711	210	hypothetical protein	
IYC_15003	49,925	50,866	942	transketolase, pyridine binding subunit	COG3958 Transketolase, C-terminal subunit
IYC_15008	50,866	51,690	825	transketolase, thiamine diphosphate binding subunit	COG3959 Transketolase, N-terminal subunit
IYC_15013	51,946	52,299	354	PemK family protein	COG2337 Growth inhibitor
IYC_15018	52,256	52,579	324	hypothetical protein	COG0864 Predicted transcriptional regulators containing the CopG/Arc/MetJ DNA-binding domain and a metal-binding domain
IYC_15023	52,718	53,878	1,161	alanine racemase	COG0787 Alanine racemase
IYC_15028	53,906	54,520	615	hypothetical protein	
IYC_15033	54,720	56,222	1,503	carbohydrate kinase family protein	COG0062 Uncharacterized conserved protein
IYC_15038	56,219	56,599	381	holo-(acyl-carrier-protein) synthase	COG0736 Phosphopantetheinyl transferase (holo-ACP synthase)
IYC_15043	56,751	57,113	363	hypothetical protein	
IYC_15048	57,509	58,834	1,326	Na+/H+ antiporter family protein	COG1757 Na+/H+ antiporter
IYC_15053	59,129	59,986	858	YihY family protein	COG1295 Predicted membrane protein
IYC_15058	60,261	60,497	237	hypothetical protein	
IYC_15063	60,503	60,709	207	hypothetical protein	COG0425 Predicted redox protein, regulator of disulfide bond formation
IYC_15068	60,696	61,784	1,089	hypothetical protein	COG2391 Predicted transporter component
IYC_15073	61,968	63,125	1,158	cysteine desulfurase family protein	COG0520 Selenocysteine lyase
IYC_15078	63,195	63,854	660	hypothetical protein	
IYC_15083	64,107	64,700	594	30S ribosomal protein S4	COG0522 Ribosomal protein S4 and related proteins
IYC_15088	65,157	66,332	1,176	sodium:dicarboxylate symporter family protein	COG1301 Na+/H+-dicarboxylate symporters
IYC_15093	66,732	67,214	483	hypothetical protein	
IYC_15098	67,361	68,152	792	putative MerR family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_15103	68,161	69,000	840	creatininase family protein	COG1402 Uncharacterized protein, putative amidase
IYC_15108	69,203	69,706	504	copper/zinc superoxide dismutase	COG2032 Cu/Zn superoxide dismutase
IYC_15113	69,974	71,800	1,827	glucosamine--fructose-6-phosphate aminotransferase, isomerizing	COG0449 Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains
IYC_15118	72,421	73,098	678	putative transaldolase	COG0176 Transaldolase
IYC_15123	73,117	73,764	648	haloacid dehalogenase, IA family protein	COG0637 Predicted phosphatase/phosphohexomutase
IYC_15128	73,788	74,573	786	sorbitol-6-phosphate dehydrogenase	COG1028 Dehydrogenases with different specificities(related to short-chain alcohol dehydrogenases)
IYC_15133	74,593	74,958	366	PTS system, glucitol/sorbitol-specific, IIA component	COG3731 Phosphotransferase system sorbitol-specificcomponent IIA
IYC_15138	75,174	76,178	1,005	PTS system glucitol/sorbitol-specific transporter subunit IIBC	COG3732 Phosphotransferase system sorbitol-specificcomponent IIBC
IYC_15143	76,191	76,736	546	PTS system glucitol/sorbitol-specific transporter subunit lic transporter subunit	COG3730 Phosphotransferase system sorbitol-specificcomponent IIC
IYC_15148	76,780	77,175	396	glucitol operon activator protein	COG4578 Glucitol operon activator
IYC_15153	77,189	78,238	1,050	putative L-iditol 2-dehydrogenase	COG1063 Threonine dehydrogenase and related Zn-dependent dehydrogenases
IYC_15158	78,273	79,244	972	putative glucitol operon regulator	COG2390 Transcriptional regulator, contains sigma factor-related N-terminal domain
IYC_15163	79,597	80,946	1,350	phosphoglucosamine mutase	COG1109 Phosphomannomutase
IYC_15168	81,081	81,611	531	2-oxoacid:acceptor oxidoreductase subunit gamma	COG1014 Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, gamma subunit
IYC_15173	81,613	82,362	750	putative 2-oxoacid:acceptor oxidoreductase subunit beta	COG1013 Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, beta subunit

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_15178	82,362	83,432	1,071	2-ketoisovalerate ferredoxin reductase	COG0674 Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, alpha subunit
IYC_15183	83,448	83,657	210	putative 2-oxoacid:acceptor oxidoreductase, delta subunit	COG1146 Ferredoxin
IYC_15188	83,759	84,421	663	hypothetical protein	COG1192 ATPases involved in chromosome partitioning
IYC_15193	84,685	85,755	1,071	butyrate kinase	COG3426 Butyrate kinase
IYC_15198	85,784	86,695	912	phosphate butyryltransferase	COG0280 Phosphotransacetylase
IYC_15203	86,688	87,782	1,095	butyrate kinase	COG3426 Butyrate kinase
IYC_15208	87,882	89,480	1,599	oxidoreductase, FAD-binding protein	COG2509 Uncharacterized FAD-dependent dehydrogenases
IYC_15213	89,482	90,768	1,287	hypothetical protein	COG4856 Uncharacterized protein conserved in bacteria
IYC_15218	90,738	91,586	849	hypothetical protein	COG1624 Uncharacterized conserved protein
IYC_15223	91,841	92,374	534	cell wall hydrolase family protein	COG3773 Cell wall hydrolyses involved in spore germination
IYC_15228	92,543	93,406	864	tRNA uridine 5-carboxymethylaminomethyl modification protein GidA	COG0492 Thioredoxin reductase
IYC_15233	93,472	93,789	318	thioredoxin	COG0526 Thiol-disulfide isomerase and thioredoxins
IYC_15238	93,898	95,073	1,176	isoaspartyl dipeptidase	COG0044 Dihydroorotase and related cyclic amidohydrolases
IYC_15243	95,148	96,401	1,254	TPR repeat-containing protein	COG0457 FOG: TPR repeat
IYC_15248	96,429	97,172	744	putative lipoprotein	
IYC_15253	97,282	98,898	1,617	phosphoenolpyruvate-protein phosphotransferase	COG1080 Phosphoenolpyruvate-protein kinase (PTS system EI component in bacteria)
IYC_15258	98,914	99,087	174	hypothetical protein	
IYC_15263	99,219	101,228	2,010	putative sensory box sigma-54 dependent transcriptional regulator	COG2202 FOG: PAS/PAC domain
IYC_15268	101,557	102,264	708	TPR repeat-containing protein	COG0457 FOG: TPR repeat
IYC_15273	102,333	102,887	555	hydro-lyase, Fe-S type, tartrate/fumarate subfamily, beta region	COG1838 Tartrate dehydratase beta subunit/Fumarate hydratase class I, C-terminal domain
IYC_15278	103,002	103,844	843	fumarate hydratase	COG1951 Tartrate dehydratase alpha subunit/Fumaratehydratase class I, N-terminal domain
IYC_15283	103,985	104,689	705	N-acetylmuramoyl-L-alanine amidase	COG0860 N-acetylmuramoyl-L-alanine amidase
IYC_15288	104,949	106,565	1,617	Na/Pi-cotransporter family protein/PhoU family protein	COG1283 Na+/phosphate symporter
IYC_15293	106,856	107,248	393	30S ribosomal protein S9	COG0103 Ribosomal protein S9
IYC_15298	107,274	107,708	435	50S ribosomal protein L13	COG0102 Ribosomal protein L13
IYC_15303	107,832	108,572	741	tRNA pseudouridine synthase A	COG0101 Pseudouridylate synthase
IYC_15308	108,572	109,375	804	cobalt ABC transporter permease	COG0619 ABC-type cobalt transport system, permease component CbiQ and related transporters
IYC_15313	109,372	110,238	867	cobalt transporter ATP-binding subunit	COG1122 ABC-type cobalt transport system, ATPase component; overlaps another CDS with the same product name
IYC_15318	110,223	111,071	849	cobalt transporter ATP-binding subunit	COG1122 ABC-type cobalt transport system, ATPase component; overlaps another CDS with the same product name
IYC_15323	111,162	111,503	342	50S ribosomal protein L17	COG0203 Ribosomal protein L17
IYC_15328	111,547	112,494	948	DNA-directed RNA polymerase subunit alpha	COG0202 DNA-directed RNA polymerase, alpha subunit/40 kD subunit
IYC_15333	112,562	113,182	621	30S ribosomal protein S4	COG0522 Ribosomal protein S4 and related proteins
IYC_15338	113,217	113,615	399	30S ribosomal protein S11	COG0100 Ribosomal protein S11
IYC_15343	113,633	114,004	372	30S ribosomal protein S13	COG0099 Ribosomal protein S13
IYC_15348	114,112	114,225	114	50S ribosomal protein L36	COG0257 Ribosomal protein L36
IYC_15353	114,242	114,460	219	translation initiation factor IF-1	COG0361 Translation initiation factor 1 (IF-1)
IYC_15358	114,468	114,749	282	hypothetical protein	COG2163 Ribosomal protein L14E/L6E/L27E
IYC_15363	114,764	115,513	750	methionine aminopeptidase, type I	COG0024 Methionine aminopeptidase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_15368	115,513	116,163	651	adenylate kinase	COG0563 Adenylate kinase and related kinases
IYC_15373	116,187	117,464	1,278	preprotein translocase subunit SecY	COG0201 Preprotein translocase subunit SecY
IYC_15378	117,465	117,905	441	50S ribosomal protein L15	COG0200 Ribosomal protein L15
IYC_15383	117,925	118,104	180	50S ribosomal protein L30	COG1841 Ribosomal protein L30/L7E
IYC_15388	118,118	118,615	498	30S ribosomal protein S5	COG0098 Ribosomal protein S5
IYC_15393	118,633	118,992	360	50S ribosomal protein L18	COG0256 Ribosomal protein L18
IYC_15398	119,011	119,553	543	50S ribosomal protein L6	COG0097 Ribosomal protein L6P/L9E
IYC_15403	119,577	119,975	399	30S ribosomal protein S8	COG0096 Ribosomal protein S8
IYC_15408	120,005	120,190	186	30S ribosomal protein S14	COG0199 Ribosomal protein S14
IYC_15413	120,207	120,749	543	50S ribosomal protein L5	COG0094 Ribosomal protein L5
IYC_15418	120,772	121,089	318	50S ribosomal protein L24	COG0198 Ribosomal protein L24
IYC_15423	121,113	121,481	369	50S ribosomal protein L14	COG0093 Ribosomal protein L14
IYC_15428	121,516	121,770	255	30S ribosomal protein S17	COG0186 Ribosomal protein S17
IYC_15433	121,792	122,004	213	50S ribosomal protein L29	COG0255 Ribosomal protein L29
IYC_15438	121,994	122,437	444	50S ribosomal protein L16	COG0197 Ribosomal protein L16/L10E
IYC_15443	122,455	123,126	672	30S ribosomal protein S3	COG0092 Ribosomal protein S3
IYC_15448	123,144	123,479	336	50S ribosomal protein L22	COG0091 Ribosomal protein L22
IYC_15453	123,501	123,785	285	30S ribosomal protein S19	COG0185 Ribosomal protein S19
IYC_15458	123,854	124,687	834	50S ribosomal protein L2	COG0090 Ribosomal protein L2
IYC_15463	124,742	125,035	294	50S ribosomal protein L23	COG0089 Ribosomal protein L23
IYC_15468	125,035	125,655	621	50S ribosomal protein L4	COG0088 Ribosomal protein L4
IYC_15473	125,680	126,309	630	50S ribosomal protein L3	COG0087 Ribosomal protein L3
IYC_15478	126,399	126,707	309	30S ribosomal protein S10	COG0051 Ribosomal protein S10
Genomic Scaffold: NZ_JH470553					
IYC_15483	<1	149	>149	hypothetical protein	
IYC_15488	146	415	270	Integrase catalytic region	COG2801 Transposase and inactivated derivatives
IYC_15493	802	2,910	2,109	sensory box histidine kinase	COG0642 Signal transduction histidine kinase
IYC_15498	3,024	3,188	165	hypothetical protein	
IYC_15503	3,516	4,580	1,065	undecaprenyldiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase	COG0707 UDP-N-acetylglucosamine:LPS N-acetylglucosamine transferase
IYC_15508	5,047	8,625	3,579	pyruvate ferredoxin oxidoreductase	COG0674 Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, alpha subunit
IYC_15513	8,890	9,006	117	hypothetical protein	
IYC_15518	9,262	9,660	399	hypothetical protein	
IYC_15523	9,819	10,460	642	HAD family hydrolase	COG0560 Phosphoserine phosphatase
IYC_15528	10,587	11,420	834	putative branched-chain amino acid aminotransferase	COG0115 Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase
IYC_15533	11,676	12,503	828	BglG family transcriptional antiterminator	COG3711 Transcriptional antiterminator
IYC_15538	12,586	14,541	1,956	PTS system, glucose family, IIABC component	COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific
IYC_15543	14,794	16,827	2,034	hypothetical protein	COG1315 Predicted polymerase, most proteins containPALM domain, HD hydrolase domain and Zn-ribbon domain
IYC_15548	16,843	17,289	447	chemotaxis protein CheW	COG0835 Chemotaxis signal transduction protein
IYC_15553	17,306	17,794	489	chemoreceptor glutamine deamidase CheD	COG1871 Chemotaxis protein; stimulates methylation of MCP proteins
IYC_15558	17,817	18,881	1,065	protein-glutamate methylesterase CheB	COG2201 Chemotaxis response regulator containing a CheY-like receiver domain and a methylesterase domain
IYC_15563	18,897	19,667	771	chemotaxis protein methyltransferase CheR	COG1352 Methylase of chemotaxis methyl-accepting proteins
IYC_15568	19,680	21,755	2,076	chemotaxis protein CheA	COG0643 Chemotaxis protein histidine kinase and related kinases
IYC_15573	21,781	22,380	600	chemotaxis protein CheC	COG1776 Chemotaxis protein CheC, inhibitor of MCP methylation
IYC_15578	22,408	22,767	360	chemotaxis protein CheY	COG0784 FOG: CheY-like receiver
IYC_15583	22,867	23,262	396	putative chemotaxis protein CheW	COG0835 Chemotaxis signal transduction protein

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_15588	23,290	24,285	996	flagellar motor switch protein FliM	COG1868 Flagellar motor switch protein
IYC_15593	24,278	25,438	1,161	flagellar motor switch protein	COG1776 Chemotaxis protein CheC, inhibitor of MCP methylation
IYC_15598	26,033	26,314	282	regulator of flagellin synthesis FlgM	COG2747 Negative regulator of flagellin synthesis (anti-sigma28 factor)
IYC_15603	26,314	26,718	405	flgN family protein	
IYC_15608	26,979	28,871	1,893	flagellar hook-associated protein FlgK	COG1256 Flagellar hook-associated protein
IYC_15613	29,017	30,174	1,158	flagellar hook-associated protein 3	COG1344 Flagellin and related hook-associated proteins
IYC_15618	30,300	30,731	432	flagellar assembly protein FliW	COG1699 Uncharacterized protein conserved in bacteria
IYC_15623	30,731	30,949	219	carbon storage regulator	COG1551 Carbon storage regulator (could also regulate swarming and quorum sensing)
IYC_15628	30,965	31,321	357	flagellar protein FlaG	COG1334 Uncharacterized flagellar protein FlaG
IYC_15633	31,378	31,677	300	hypothetical protein	
IYC_15638	31,714	32,100	387	flagellar protein FliS	COG1516 Flagellin-specific chaperone FliS
IYC_15643	32,192	34,645	2,454	flagellar hook-associated protein 2	COG1345 Flagellar capping protein
IYC_15648	34,654	34,995	342	hypothetical protein	
IYC_15653	35,236	36,066	831	flagellin	COG1344 Flagellin and related hook-associated proteins
IYC_15658	36,233	38,347	2,115	glycosyltransferase	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_15663	38,370	39,479	1,110	glycosyl transferase family protein	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_15668	39,479	39,841	363	hypothetical protein	
IYC_15673	39,895	41,937	2,043	hypothetical protein	COG0457 FOG: TPR repeat
IYC_15678	42,005	43,417	1,413	glycosyl transferase family 2	COG0457 FOG: TPR repeat
IYC_15683	43,455	44,234	780	glucose-1-phosphate cytidylyltransferase	COG1208 Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysaccharide biosynthesis/translation initiation factor 2B, gamma/epsilon subunits (eIF-2Bgamma/eIF-2Bepsilon)
IYC_15688	44,262	45,353	1,092	hypothetical protein	COG0451 Nucleoside-diphosphate-sugar epimerases
IYC_15693	45,403	46,659	1,257	putative UDP-glucose 6-dehydrogenase	COG1004 Predicted UDP-glucose 6-dehydrogenase
IYC_15698	46,670	47,977	1,308	putative glycosyl transferase family 2	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_15703	48,008	49,861	1,854	capsule polysaccharide biosynthesis protein	COG1086 Predicted nucleoside-diphosphate sugar epimerases
IYC_15708	49,875	50,822	948	glycosyl transferase family protein	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_15713	50,892	51,524	633	methyltransferase type 11	COG0500 SAM-dependent methyltransferases
IYC_15718	51,672	51,818	147	hypothetical protein	
IYC_15723	51,815	51,991	177	hypothetical protein	
IYC_15728	52,055	52,393	339	pseudaminic acid biosynthesis protein PseA	
IYC_15733	52,610	53,614	1,005	UDP-4-dehydro-6-deoxy-2-acetamido-D-glucose 4-reductase	COG1086 Predicted nucleoside-diphosphate sugar epimerases
IYC_15738	53,708	54,442	735	putative polysaccharide biosynthesis protein with acetyltransferase domain	COG1861 Spore coat polysaccharide biosynthesis protein F, CMP-KDO synthetase homolog
IYC_15743	54,601	55,701	1,101	spore coat polysaccharide biosynthesis protein SpsE	COG2089 Sialic acid synthase
IYC_15748	55,703	56,614	912	methionyl-tRNA formyltransferase	COG0223 Methionyl-tRNA formyltransferase
IYC_15753	56,607	57,698	1,092	putative uDP-4-amino-4-deoxy-L-arabinose--oxoglutarate aminotransferase	COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis
IYC_15758	57,749	58,282	534	GCN5-related N-acetyltransferase	COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins
IYC_15763	58,364	59,089	726	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_15768	59,215	59,895	681	LmbE-like protein	COG2120 Uncharacterized proteins, LmbE homologs
IYC_15773	59,905	60,915	1,011	glycosyl transferase-like protein	COG3980 Spore coat polysaccharide biosynthesis protein, predicted glycosyltransferase
IYC_15778	60,950	62,161	1,212	hypothetical protein	
IYC_15783	62,197	63,243	1,047	putative glycosyltransferase	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_15788	63,514	64,335	822	flagellin	COG1344 Flagellin and related hook-associated proteins
IYC_15793	64,377	65,225	849	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_15798	65,567	65,734	168	hypothetical protein	
IYC_15803	65,853	66,026	174	hypothetical protein	
IYC_15808	66,156	66,332	177	hypothetical protein	
IYC_15813	66,748	67,710	963	cell envelope-like transcriptional attenuator domain-containing protein	COG1316 Transcriptional regulator
IYC_15818	67,749	68,426	678	capsular polysaccharide biosynthsis protein	COG3944 Capsular polysaccharide biosynthesis protein
IYC_15823	68,438	69,211	774	exopolysaccharide biosynthesis protein	COG4464 Capsular polysaccharide biosynthesis protein
IYC_15828	69,230	69,940	711	capsular exopolysaccharide family protein	COG0489 ATPases involved in chromosome partitioning
IYC_15833	70,067	70,735	669	capsular polysaccharide biosynthsis protein	COG2148 Sugar transferases involved in lipopolysaccharide synthesis
IYC_15838	70,813	72,135	1,323	nucleotide sugar dehydrogenase	COG0677 UDP-N-acetyl-D-mannosaminuronate dehydrogenase
IYC_15843	72,204	73,472	1,269	hypothetical protein	
IYC_15848	73,560	74,606	1,047	glycosytransferase	COG0438 Glycosyltransferase
IYC_15853	74,658	76,208	1,551	integral membrane protein MviN	COG0728 Uncharacterized membrane protein, putative virulence factor
IYC_15858	76,223	76,651	429	serine acetyltransferase	COG1045 Serine acetyltransferase
IYC_15863	76,676	77,872	1,197	hypothetical protein	
IYC_15868	77,875	78,483	609	hypothetical protein	
IYC_15873	78,517	79,158	642	galactoside acetyltransferase LacA	COG0110 Acetyltransferase (isoleucine patch superfamily)
IYC_15878	79,175	80,257	1,083	hypothetical protein	
IYC_15883	80,278	81,606	1,329	coenzyme F390 synthetase	COG1541 Coenzyme F390 synthetase
IYC_15888	81,609	83,561	1,953	heparinase II/III-like protein	
IYC_15893	83,548	84,288	741	beta-1,4-N-acetyl-mannosaminyltransferase	COG1922 Teichoic acid biosynthesis proteins
IYC_15898	84,565	85,014	450	VanZ family protein	COG5652 Predicted integral membrane protein
IYC_15903	85,064	85,891	828	hypothetical protein	COG2755 Lysophospholipase L1 and related esterases
IYC_15908	86,338	86,523	186	hypothetical protein	
IYC_15913	86,897	87,760	864	hypothetical protein	COG4129 Predicted membrane protein
IYC_15918	88,088	89,518	1,431	cardiolipin synthetase	COG1502 Phosphatidylserine/phosphatidylglycerophosphate/cardiolipi n synthases and related enzymes
IYC_15923	89,697	90,041	345	hypothetical protein	COG1344 Flagellin and related hook-associated proteins
IYC_15928	90,064	90,402	339	hypothetical protein	
IYC_15933	90,497	90,682	186	hypothetical protein	
IYC_15938	90,696	91,889	1,194	argininosuccinate synthase	COG0137 Argininosuccinate synthase
IYC_15943	91,911	93,233	1,323	argininosuccinate lyase	COG0165 Argininosuccinate lyase
IYC_15948	93,668	94,036	369	hypothetical protein	
IYC_15953	94,178	95,032	855	flagellin	COG1344 Flagellin and related hook-associated proteins
IYC_15958	95,602	96,003	402	flagellar basal body rod protein FlgB	COG1815 Flagellar basal body protein
IYC_15963	96,145	96,579	435	flagellar basal body rod protein FlgC	COG1558 Flagellar basal body rod protein
IYC_15968	96,591	96,905	315	flagellar hook-basal body protein FliE	COG1677 Flagellar hook-basal body protein
IYC_15973	96,921	98,501	1,581	flagellar MS-ring protein	COG1766 Flagellar biosynthesis/type III secretory pathway lipoprotein

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_15978	98,508	99,521	1,014	flagellar motor switch protein G	COG1536 Flagellar motor switch protein
IYC_15983	99,505	100,269	765	flagellar assembly protein FliH	COG1317 Flagellar biosynthesis/type III secretory pathway protein
IYC_15988	100,315	101,625	1,311	flagellar protein export ATPase FliI	COG1157 Flagellar biosynthesis/type III secretory pathway ATPase
IYC_15993	101,714	102,151	438	flagellar protein FliJ	COG2882 Flagellar biosynthesis chaperone
IYC_15998	102,157	103,533	1,377	flagellar hook-length control protein	COG3144 Flagellar hook-length control protein
IYC_16003	103,563	104,402	840	flagellar hook capping protein	COG1843 Flagellar hook capping protein
IYC_16008	104,417	104,836	420	flagellar operon protein	
IYC_16013	104,918	105,952	1,035	flagellar hook protein flgE	COG1749 Flagellar hook protein FlgE
IYC_16018	106,224	106,433	210	flagellar protein FlbD	COG1582 Uncharacterized protein, possibly involved in motility
IYC_16023	106,430	107,242	813	chemotaxis MotA protein	COG1291 Flagellar motor component
IYC_16028	107,235	107,990	756	chemotaxis motB protein	COG1360 Flagellar motor protein
IYC_16033	107,987	108,508	522	flagellar basal body-associated protein FliL	COG1580 Flagellar basal body-associated protein
IYC_16038	108,579	108,977	399	flagellar biosynthesis domain-containing protein	
IYC_16043	108,961	109,737	777	flagellar biosynthesis protein FliP	COG1338 Flagellar biosynthesis pathway, component FliP
IYC_16048	109,749	110,018	270	flagellar biosynthesis protein FliQ	COG1987 Flagellar biosynthesis pathway, component FliQ
IYC_16053	110,052	111,881	1,830	bifunctional flagellar biosynthesis protein FliR/FlhB	COG1684 Flagellar biosynthesis pathway, component FliR
IYC_16058	111,909	113,975	2,067	flagellar biosynthesis protein FlhA	COG1298 Flagellar biosynthesis pathway, component FlhA
IYC_16063	113,972	115,219	1,248	flagellar biosynthesis regulator FlhF	COG1419 Flagellar GTP-binding protein
IYC_16068	115,213	116,073	861	flagellar biosynthesis protein FlhG	COG0455 ATPases involved in chromosome partitioning
IYC_16073	116,088	116,747	660	hypothetical protein	COG5581 Predicted glycosyltransferase
IYC_16078	116,761	117,489	729	RNA polymerase sigma factor for flagellar operonFliA	COG1191 DNA-directed RNA polymerase specialized sigma subunit
IYC_16083	117,516	118,073	558	hypothetical protein	
IYC_16088	118,151	118,267	117	hypothetical protein	
IYC_16093	118,333	119,100	768	flagellar basal body rod protein FlgG	COG4786 Flagellar basal body rod protein
IYC_16098	119,147	119,938	792	flagellar basal body rod protein FlgG	COG4786 Flagellar basal body rod protein
IYC_16103	120,322	120,897	576	hypothetical protein	
IYC_16108	121,350	121,862	513	hypothetical protein	
IYC_16113	123,007	124,650	1,644	putative manganese-dependent inorganic pyrophosphatase	COG1227 Inorganic pyrophosphatase/exopolyphosphatase
IYC_16118	125,195	127,741	2,547	putative calcium-translocating P-type ATPase, PMCA-type	COG0474 Cation transport ATPase
IYC_16123	127,780	128,979	1,200	hypothetical protein	COG2133 Glucose/sorbosone dehydrogenases
IYC_16128	129,002	129,721	720	putative phospholipase C	
IYC_16133	129,894	130,469	576	hypothetical protein	COG2226 Methylase involved in ubiquinone/menaquinone biosynthesis
IYC_16138	130,444	132,123	1,680	putative B12-binding Fe-S oxidoreductase	COG1032 Fe-S oxidoreductase
IYC_16143	132,282	132,608	327	putative V-type ATPase, G subunit	COG3599 Cell division initiation protein
IYC_16148	132,595	134,553	1,959	V-type ATP synthase subunit I	COG1269 Archaeal/vacuolar-type H+-ATPase subunit I
IYC_16153	134,578	135,066	489	V-type ATP synthase subunit K	COG0636 F0F1-type ATP synthase, subunit c/Archaeal/vacuolar-type H+-ATPase, subunit K
IYC_16158	135,082	135,681	600	V-type ATPase, E subunit	COG1390 Archaeal/vacuolar-type H+-ATPase subunit E
IYC_16163	135,694	136,695	1,002	V-type ATP synthase subunit C	COG1527 Archaeal/vacuolar-type H+-ATPase subunit C
IYC_16168	136,688	136,996	309	V-type ATP synthase subunit F	COG1436 Archaeal/vacuolar-type H+-ATPase subunit F

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_16173	137,016	138,788	1,773	V-type ATP synthase subunit A	COG1155 Archaeal/vacuolar-type H+-ATPase subunit A
IYC_16178	138,781	140,166	1,386	V-type ATP synthase subunit B	COG1156 Archaeal/vacuolar-type H+-ATPase subunit B
IYC_16183	140,195	140,845	651	V-type ATP synthase subunit D	COG1394 Archaeal/vacuolar-type H+-ATPase subunit D
IYC_16188	141,059	141,595	537	peptide deformylase	COG0242 N-formylmethionyl-tRNA deformylase
IYC_16193	141,613	142,476	864	nucleoside hydrolase, IUNH family protein	COG1957 Inosine-uridine nucleoside N-ribohydrolase
IYC_16198	142,602	142,892	291	hypothetical protein	
IYC_16203	143,311	143,403	93	hypothetical protein	
IYC_16208	143,702	144,208	507	hypothetical protein	
IYC_16213	144,527	145,696	1,170	metallo-beta-lactamase family protein/flavodoxin	COG0426 Uncharacterized flavoproteins
IYC_16218	146,161	146,556	396	hypothetical protein	
IYC_16223	146,591	146,764	174	hypothetical protein	
IYC_16228	147,248	148,222	975	fructose 1,6-bisphosphatase II	COG1494 Fructose-1,6-bisphosphatase/sedoheptulose 1,7-bisphosphatase and related proteins
IYC_16233	148,450	149,424	975	putative thiosulfate sulfurtransferase	COG2897 Rhodanese-related sulfurtransferase
IYC_16238	149,608	150,282	675	group 2 family glycosyl transferase protein	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_16243	150,282	150,983	702	hypothetical protein	COG3222 Uncharacterized protein conserved in bacteria
IYC_16248	150,961	151,959	999	hypothetical protein	COG0165 Argininosuccinate lyase
IYC_16253	151,959	153,104	1,146	cysteine-rich domain-containing protein	COG0247 Fe-S oxidoreductase
IYC_16258	153,213	153,932	720	DedA family protein	COG0398 Uncharacterized conserved protein; overlapsanother CDS with the same product name
IYC_16263	153,898	154,626	729	DedA family protein	COG0398 Uncharacterized conserved protein; overlapsanother CDS with the same product name
IYC_16268	154,897	155,085	189	hypothetical protein	
IYC_16273	155,117	156,040	924	HPr kinase/phosphorylase	COG1493 Serine kinase of the HPr protein, regulatescarbohydrate metabolism
IYC_16278	156,033	156,587	555	5-formyltetrahydrofolate cyclo-ligase family protein	COG0212 5-formyltetrahydrofolate cyclo-ligase
IYC_16283	156,832	157,767	936	putative magnesium and cobalt transport protein CorA	COG0598 Mg2+ and Co2+ transporters
IYC_16288	157,758	159,293	1,536	putative aminopeptidase 1	COG1362 Aspartyl aminopeptidase
IYC_16293	159,321	159,782	462	hypothetical protein	COG0613 Predicted metal-dependent phosphoesterases (PHP family)
IYC_16298	160,351	161,406	1,056	hypothetical protein	COG0655 Multimeric flavodoxin WrbA
IYC_16303	161,564	162,787	1,224	hypothetical protein	COG0039 Malate/lactate dehydrogenases
IYC_16308	162,864	167,729	4,866	hypothetical protein	COG1057 Nicotinic acid mononucleotide adenylyltransferase
IYC_16313	168,144	168,620	477	putative lipoprotein	
IYC_16318	168,734	169,558	825	HAD superfamily hydrolase	COG0561 Predicted hydrolases of the HAD superfamily
IYC_16323	169,754	170,023	270	hypothetical protein	COG1937 Uncharacterized protein conserved in bacteria
IYC_16328	170,175	170,663	489	hypothetical protein	COG3610 Uncharacterized conserved protein
IYC_16333	170,660	171,430	771	hypothetical protein	COG2966 Uncharacterized conserved protein
IYC_16338	171,605	172,846	1,242	serine hydroxymethyltransferase	COG0112 Glycine/serine hydroxymethyltransferase
IYC_16343	173,409	174,353	945	putative amino acid kinase	COG0549 Carbamate kinase
IYC_16348	174,455	175,456	1,002	ornithine carbamoyltransferase	COG0078 Ornithine carbamoyltransferase
IYC_16353	175,703	179,464	3,762	phosphoribosylformylglycinamidine synthase	COG0046 Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain
IYC_16358	179,822	180,298	477	copper amine oxidase domain-containing protein	
IYC_16363	180,438	181,091	654	hypothetical protein	
IYC_16368	181,470	182,804	1,335	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_16373	183,568	184,935	1,368	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter
IYC_16378	185,057	186,046	990	ornithine cyclodeaminase	COG2423 Predicted ornithine cyclodeaminase, mu-crystallin homolog
IYC_16383	186,290	186,874	585	hypothetical protein	COG0425 Predicted redox protein, regulator of disulfide bond formation
IYC_16388	187,223	188,092	870	DegV family protein	COG1307 Uncharacterized protein conserved in bacteria
IYC_16393	188,196	189,104	909	auxin efflux carrier family protein	COG0679 Predicted permeases
IYC_16398	189,238	190,005	768	histidinol-phosphatase	COG1387 Histidinol phosphatase and related hydrolases of the PHP family
IYC_16403	190,181	191,428	1,248	recombination factor protein RarA	COG2256 ATPase related to the helicase subunit of the Holliday junction resolvase
IYC_16408	191,592	192,038	447	rrF2 family protein	COG1959 Predicted transcriptional regulator
IYC_16413	192,031	193,224	1,194	cysteine desulfurase	COG1104 Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes
IYC_16418	193,226	193,654	429	putative iron-sulfur cluster assembly protein	COG0822 NifU homolog involved in Fe-S cluster formation
IYC_16423	194,106	194,594	489	PRC-barrel domain-containing protein	COG3881 Uncharacterized protein conserved in bacteria
IYC_16428	194,607	195,635	1,029	hypothetical protein	COG0628 Predicted permease
IYC_16433	195,956	198,595	2,640	alanyl-tRNA synthetase	COG0013 Alanyl-tRNA synthetase
IYC_16438	198,711	198,962	252	hypothetical protein	COG4472 Uncharacterized protein conserved in bacteria
IYC_16443	199,233	199,646	414	Holliday junction resolvase-like protein	COG0816 Predicted endonuclease involved in recombination (possible Holliday junction resolvase in Mycoplasmas and B. subtilis)
IYC_16448	199,660	199,914	255	hypothetical protein	
IYC_16453	200,034	200,486	453	Fur family transcriptional regulator	COG0735 Fe2+/Zn2+ uptake regulation proteins
IYC_16458	200,534	202,243	1,710	RNA-metabolizing metallo-beta-lactamase family protein	COG0595 Predicted hydrolase of the metallo-beta-lactamase superfamily
IYC_16463	202,458	202,574	117	hypothetical protein	
IYC_16468	202,890	204,716	1,827	GTP-binding protein TypA	COG1217 Predicted membrane GTPase involved in stress response
IYC_16473	204,793	205,824	1,032	hypothetical protein	COG1559 Predicted periplasmic solute-binding protein
IYC_16478	205,907	206,563	657	O-methyltransferase family protein	COG4122 Predicted O-methyltransferase
IYC_16483	206,556	207,782	1,227	peptidase, U32 family protein	COG0826 Collagenase and related proteases
IYC_16488	207,931	208,551	621	uridine/cytidine kinase	COG0572 Uridine kinase
IYC_16493	208,637	210,307	1,671	penicillin-binding protein	COG0768 Cell division protein FtsI/penicillin-binding protein 2
IYC_16498	210,459	211,541	1,083	phage integrase family site specific recombinase	COG0582 Integrase
IYC_16503	211,593	212,039	447	hypothetical protein	COG2856 Predicted Zn peptidase
IYC_16508	212,058	212,486	429	immunity repressor protein	COG1396 Predicted transcriptional regulators
IYC_16513	212,721	212,942	222	hypothetical protein	COG1476 Predicted transcriptional regulators
IYC_16518	213,004	213,303	300	hypothetical protein	
IYC_16523	213,345	213,527	183	hypothetical protein	
IYC_16528	213,544	213,810	267	phage protein	
IYC_16533	213,859	214,218	360	hypothetical protein	
IYC_16538	214,215	214,367	153	hypothetical protein	
IYC_16543	214,370	216,304	1,935	hypothetical protein	COG1196 Chromosome segregation ATPases
IYC_16548	216,306	216,701	396	hypothetical protein	
IYC_16553	216,702	217,448	747	phage recombination protein Bet	COG0208 Ribonucleotide reductase, beta subunit
IYC_16558	217,448	218,161	714	hypothetical protein	COG1235 Metal-dependent hydrolases of the beta-lactamase superfamily I
IYC_16563	218,176	218,796	621	hypothetical protein	COG3481 Predicted HD-superfamily hydrolase
IYC_16568	218,800	219,207	408	single-strand binding protein family	COG0629 Single-stranded DNA-binding protein
IYC_16573	219,264	219,467	204	hypothetical protein	
IYC_16578	219,554	219,721	168	hypothetical protein	
IYC_16583	219,775	219,942	168	hypothetical protein	
IYC_16588	220,017	221,012	996	phage integrase	COG0582 Integrase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_16593	221,026	221,319	294	hypothetical protein	
IYC_16598	221,366	221,692	327	hypothetical protein	
IYC_16603	221,782	222,045	264	hypothetical protein	
IYC_16608	222,048	222,239	192	hypothetical protein	
IYC_16613	222,236	222,466	231	hypothetical protein	
IYC_16618	222,468	223,001	534	hypothetical protein	
IYC_16623	223,254	223,406	153	hypothetical protein	
IYC_16628	223,513	223,722	210	hypothetical protein	
IYC_16633	223,723	224,055	333	MazG nucleotide pyrophosphohydrolase	COG1694 Predicted pyrophosphatase
IYC_16638	224,058	224,642	585	hypothetical protein	
IYC_16643	224,675	227,269	2,595	DNA primase related protein	COG0358 DNA primase (bacterial type)
IYC_16648	227,581	228,357	777	hypothetical protein	COG3561 Phage anti-repressor protein
IYC_16653	228,369	228,812	444	hypothetical protein	
IYC_16658	228,968	229,384	417	hypothetical protein	COG3335 Transposase and inactivated derivatives
IYC_16663	229,646	229,837	192	hypothetical protein	
IYC_16668	229,890	230,741	852	hypothetical protein	
IYC_16673	230,757	232,490	1,734	hypothetical protein	
IYC_16678	232,494	233,762	1,269	SPP1 family phage portal protein	
IYC_16683	233,762	234,880	1,119	hypothetical protein	COG5585 NAD+--asparagine ADP-ribosyltransferase
IYC_16688	234,883	235,050	168	hypothetical protein	
IYC_16693	235,186	235,779	594	hypothetical protein	
IYC_16698	235,798	236,688	891	major head protein	
IYC_16703	236,734	236,904	171	hypothetical protein	
IYC_16708	236,913	237,176	264	hypothetical protein	
IYC_16713	237,178	237,588	411	hypothetical protein	
IYC_16718	237,588	237,941	354	hypothetical protein	
IYC_16723	237,938	238,369	432	hypothetical protein	
IYC_16728	238,370	238,870	501	hypothetical protein	
IYC_16733	238,870	239,214	345	hypothetical protein	
IYC_16738	239,214	>239446	>233	putative protein gp15	
Genomic Scaffold: NZ_JH470554					
IYC_16743	<1	239	>239	sodium:alanine symporter family protein	COG1115 Na+/alanine symporter
IYC_16748	435	1,706	1,272	hypothetical protein	COG0840 Methyl-accepting chemotaxis protein
IYC_16753	1,973	2,818	846	GGDEF domain-containing protein	COG2199 FOG: GGDEF domain
IYC_16758	2,811	4,475	1,665	STE like transcription factor domain-containing protein	COG5298 Uncharacterized protein conserved in bacteria
IYC_16763	4,481	5,743	1,263	group 2 family glycosyl transferase	COG1215 Glycosyltransferases, probably involved in cell wall biogenesis
IYC_16768	5,899	7,992	2,094	cellulose synthase domain-containing protein	
IYC_16773	8,136	9,230	1,095	hypothetical protein	
IYC_16778	9,459	9,623	165	hypothetical protein	
IYC_16793	10,832	11,398	567	hypothetical protein	
IYC_16798	11,626	11,892	267	hypothetical protein	COG1937 Uncharacterized protein conserved in bacteria
IYC_16803	11,929	14,364	2,436	copper-translocating P-type ATPase	COG2217 Cation transport ATPase
IYC_16808	14,483	14,698	216	copper chaperone CopZ	COG2608 Copper chaperone
IYC_16813	15,100	17,472	2,373	cation-transporting ATPase, P-type	COG0474 Cation transport ATPase
IYC_16818	17,762	19,480	1,719	ABC transporter ATP-binding protein/permease	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_16823	19,778	20,572	795	CAAX amino terminal protease family protein	COG1266 Predicted metal-dependent membrane protease
IYC_16828	20,733	21,089	357	dinitrogenase iron-molybdenum cofactor family protein	COG1433 Uncharacterized conserved protein
IYC_16833	21,402	22,295	894	DegV family protein	COG1307 Uncharacterized protein conserved in bacteria
IYC_16838	22,337	22,801	465	5-nitroimidazole antibiotic resistance protein	COG3467 Predicted flavin-nucleotide-binding protein
IYC_16843	23,122	23,337	216	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_16848	23,375	23,803	429	HIT family protein	COG0537 Diadenosine tetraphosphate (Ap4A) hydrolaseand other HIT family hydrolases
IYC_16853	23,959	24,333	375	hypothetical protein	
IYC_16858	24,620	25,378	759	type IV leader peptidase family protein	COG1989 Type II secretory pathway, prepilin signal peptidase PulO and related peptidases
IYC_16863	25,666	25,779	114	hypothetical protein	
IYC_16868	26,083	27,210	1,128	hypothetical protein	COG5438 Predicted multitransmembrane protein
IYC_16873	27,634	27,951	318	response regulator, LytTR family protein	COG3279 Response regulator of the LytR/AlgR family
IYC_16878	27,953	28,381	429	hypothetical protein	
IYC_16883	28,812	30,179	1,368	potassium transporter peripheral membrane component	COG0569 K+ transport systems, NAD-binding component
IYC_16888	30,176	31,621	1,446	Trk system potassium uptake protein TrkH	COG0168 Trk-type K+ transport systems, membrane components
IYC_16893	31,817	33,133	1,317	sodium- and chloride-dependent transporter	COG0733 Na+-dependent transporters of the SNF family
IYC_16898	33,725	36,292	2,568	helicase, UvrD/REP/exonuclease family protein	COG0210 Superfamily I DNA and RNA helicases
IYC_16903	38,248	39,378	1,131	hypothetical protein	COG4194 Predicted membrane protein
IYC_16908	39,393	39,767	375	GntR family transcriptional regulator	COG1725 Predicted transcriptional regulators
IYC_16913	40,246	41,034	789	putative cell wall hydrolase	COG1388 FOG: LysM repeat
IYC_16918	41,785	42,372	588	rubrerythrin	COG1592 Rubrerythrin
IYC_16923	42,668	44,128	1,461	N-acetylmuramoyl-L-alanine amidase	COG2247 Putative cell wall-binding domain
IYC_16928	44,430	46,460	2,031	PTS system glucose-specific transporter subunit IIABC	COG1263 Phosphotransferase system IIC components, glucose/maltose/N-acetylglucosamine-specific
IYC_16933	46,913	47,560	648	undecaprenyl pyrophosphate synthetase-like protein	COG0020 Undecaprenyl pyrophosphate synthase
IYC_16938	47,728	48,273	546	rubrerythrin	COG1633 Uncharacterized conserved protein
IYC_16943	48,369	48,836	468	hypothetical protein	
IYC_16948	49,142	49,567	426	putative transcriptional regulator	COG1959 Predicted transcriptional regulator
IYC_16953	49,764	51,650	1,887	carbon monoxide dehydrogenase	COG1151 6Fe-6S prismane cluster-containing protein
IYC_16958	51,794	52,237	444	nitrate reductase, iron-sulfur subunit	COG1142 Fe-S-cluster-containing hydrogenase components 2
IYC_16963	52,425	53,651	1,227	nitrate reductase, NADH oxidase subunit	COG0446 Uncharacterized NAD(FAD)-dependent dehydrogenases
Genomic Scaffold: NZ_JH470555					
IYC_16968	237	2,717	2,481	DNA gyrase subunit A	COG0188 Type IIA topoisomerase (DNA gyrase/topo II,topoisomerase IV), A subunit
IYC_16973	2,743	4,656	1,914	DNA gyrase subunit B	COG0187 Type IIA topoisomerase (DNA gyrase/topo II,topoisomerase IV), B subunit
IYC_16978	4,674	4,937	264	hypothetical protein	
IYC_16983	4,958	6,052	1,095	recombination protein F	COG1195 Recombinational DNA repair ATPase (RecF pathway)
IYC_16988	6,121	6,327	207	S4 domain-containing protein	COG2501 Uncharacterized conserved protein
IYC_16993	6,346	7,449	1,104	DNA polymerase III subunit beta	COG0592 DNA polymerase sliding clamp subunit (PCNA homolog)
IYC_16998	7,712	9,058	1,347	chromosomal replication initiation protein	COG0593 ATPase involved in DNA replication initiation
IYC_17003	9,612	9,746	135	50S ribosomal protein L34	COG0230 Ribosomal protein L34
IYC_17008	9,801	10,136	336	ribonuclease P	COG0594 RNase P protein component
IYC_17013	10,133	10,342	210	hypothetical protein	COG0759 Uncharacterized conserved protein
IYC_17018	10,368	11,021	654	putative inner membrane protein translocase component YidC	COG0706 Preprotein translocase subunit YidC
IYC_17023	11,039	11,665	627	DNA/RNA-binding protein	COG1847 Predicted RNA-binding protein
IYC_17028	11,713	13,098	1,386	tRNA modification GTPase TrmE	COG0486 Predicted GTPase
IYC_17033	13,107	14,984	1,878	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA	COG0445 NAD/FAD-utilizing enzyme apparently involved in cell division

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_17038	15,008	15,727	720	16S rRNA methyltransferase GidB	COG0357 Predicted S-adenosylmethionine-dependent methyltransferase involved in bacterial cell division
IYC_17043	15,827	16,606	780	parB family protein	COG1475 Predicted transcriptional regulators
IYC_17048	16,729	17,493	765	sporulation initiation inhibitor protein soj	COG1192 ATPases involved in chromosome partitioning
IYC_17053	17,509	18,357	849	stage 0 sporulation protein J	COG1475 Predicted transcriptional regulators
IYC_17058	18,395	18,904	510	hypothetical protein	
IYC_17063	18,957	19,202	246	hypothetical protein	
IYC_17068	19,204	19,785	582	putative sporulation protein	
IYC_17073	19,843	20,898	1,056	sporulation integral membrane protein YtvI	COG0628 Predicted permease
IYC_17078	20,912	22,066	1,155	cysteine desulfurase family protein	COG0520 Selenocysteine lyase
IYC_17083	22,257	22,502	246	hypothetical protein	
IYC_17088	22,509	23,378	870	mechanosensitive ion channel family protein	COG0668 Small-conductance mechanosensitive channel
IYC_17093	23,394	23,591	198	hypothetical protein	COG4481 Uncharacterized protein conserved in bacteria
IYC_17098	23,913	25,061	1,149	2-hydroxyglutaryl-CoA dehydratase subunit D	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB
IYC_17103	25,098	26,246	1,149	2-hydroxyglutaryl-CoA dehydratase subunit D	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB
IYC_17108	26,352	26,636	285	30S ribosomal protein S6	COG0360 Ribosomal protein S6
IYC_17113	26,652	27,098	447	single-strand binding protein	COG0629 Single-stranded DNA-binding protein
IYC_17118	27,117	27,359	243	30S ribosomal protein S18	COG0238 Ribosomal protein S18
IYC_17123	27,506	27,823	318	hypothetical protein	
IYC_17128	27,835	28,842	1,008	hypothetical protein	COG4241 Predicted membrane protein
IYC_17133	28,843	30,828	1,986	DHH family protein	COG3887 Predicted signaling protein consisting of amodified GGDEF domain and a DHH domain
IYC_17138	30,828	31,271	444	50S ribosomal protein L9	COG0359 Ribosomal protein L9
IYC_17143	31,282	33,201	1,920	ATP-dependent protease	COG1067 Predicted ATP-dependent protease
IYC_17148	33,232	34,566	1,335	replicative DNA helicase	COG0305 Replicative DNA helicase
IYC_17153	34,857	36,578	1,722	phosphoglucomutase/phosphomannomutase family protein	COG1109 Phosphomannomutase
IYC_17158	36,772	36,987	216	hypothetical protein	
IYC_17163	37,157	38,542	1,386	pyridine nucleotide-disulfide oxidoreductase family protein	COG2509 Uncharacterized FAD-dependent dehydrogenases
IYC_17168	38,776	40,062	1,287	adenylosuccinate synthetase	COG0104 Adenylosuccinate synthase
IYC_17173	40,185	40,370	186	hypothetical protein	
IYC_17178	40,475	41,224	750	acyl-ACP thioesterase family protein	COG3884 Acyl-ACP thioesterase
IYC_17183	41,325	42,554	1,230	pyridine nucleotide-disulfide oxidoreductase family protein	COG2081 Predicted flavoproteins
IYC_17188	42,735	43,109	375	CoA-binding protein	COG1832 Predicted CoA-binding protein
IYC_17193	43,122	43,931	810	PEP synthetase regulatory protein	COG1806 Uncharacterized protein conserved in bacteria
IYC_17198	44,147	44,902	756	dihydrodipicolinate reductase	COG0289 Dihydrodipicolinate reductase
IYC_17203	45,181	46,059	879	dihydrodipicolinate synthase	COG0329 Dihydrodipicolinate synthase/N-acetylneuraminate lyase
IYC_17208	46,082	47,074	993	aspartate-semialdehyde dehydrogenase	COG0136 Aspartate-semialdehyde dehydrogenase
IYC_17213	47,560	49,242	1,683	oligopeptide/dipeptide ABC transporter, oligopeptide/dipeptide-binding protein	COG4166 ABC-type oligopeptide transport system, periplasmic component
IYC_17218	49,323	50,252	930	oligopeptide/dipeptide ABC transporter, permeaseprotein	COG0601 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components
IYC_17223	50,268	51,185	918	oligopeptide/dipeptide ABC transporter, permeaseprotein	COG1173 ABC-type dipeptide/oligopeptide/nickel transport systems, permease components

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_17228	51,200	52,219	1,020	oligopeptide/dipeptide ABC transporter, ATP-binding protein	COG0444 ABC-type dipeptide/oligopeptide/nickel transport system, ATPase component
IYC_17233	52,219	53,178	960	oligopeptide/dipeptide ABC transporter, ATP-binding protein	COG4608 ABC-type oligopeptide transport system, ATPase component
IYC_17238	53,415	53,687	273	small acid-soluble spore protein	
IYC_17243	53,773	54,513	741	hypothetical protein	COG0500 SAM-dependent methyltransferases
IYC_17248	54,551	55,441	891	Hsp33-like chaperonin	COG1281 Disulfide bond chaperones of the HSP33 family
IYC_17258	56,405	57,100	696	ABC transporter, ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_17263	57,102	58,673	1,572	ABC transporter permease	
IYC_17268	58,894	59,601	708	hypothetical protein	
IYC_17273	59,756	60,655	900	hypothetical protein	
IYC_17278	61,003	62,910	1,908	threonyl-tRNA synthetase	COG0441 Threonyl-tRNA synthetase
IYC_17283	63,164	63,718	555	translation initiation factor IF-3	COG0290 Translation initiation factor 3 (IF-3)
IYC_17288	63,740	63,937	198	50S ribosomal protein L35	COG0291 Ribosomal protein L35
IYC_17293	63,966	64,325	360	50S ribosomal protein L20	COG0292 Ribosomal protein L20
IYC_17298	64,447	65,790	1,344	TrkH family potassium uptake protein	COG0168 Trk-type K+ transport systems, membrane components
IYC_17303	65,801	66,463	663	TrkA family potassium uptake protein	COG0569 K+ transport systems, NAD-binding component
IYC_17308	66,524	67,303	780	RNA methyltransferase, TrmH family	COG0566 rRNA methylases
IYC_17313	67,648	68,667	1,020	phenylalanyl-tRNA synthetase subunit alpha	COG0016 Phenylalanyl-tRNA synthetase alpha subunit
IYC_17318	68,829	71,210	2,382	phenylalanyl-tRNA synthetase subunit beta	COG0073 EMAP domain
IYC_17323	71,412	72,008	597	hypothetical protein	COG3027 Uncharacterized protein conserved in bacteria
IYC_17328	72,513	73,262	750	exosporium protein	
IYC_17333	73,925	74,902	978	DNA replication protein DnaC	COG1484 DNA replication protein
IYC_17338	74,895	75,881	987	putative DNA replication protein DnaD	COG3935 Putative primosome component and related proteins
IYC_17343	76,062	76,997	936	putative peptidase	COG0739 Membrane proteins related to metalloendopeptidases
IYC_17348	77,095	78,153	1,059	2-nitropropane dioxygenase family oxidoreductase	COG2070 Dioxygenases related to 2-nitropropane dioxygenase
IYC_17353	78,169	78,645	477	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_17358	78,638	79,618	981	3-oxoacyl-(acyl-carrier-protein) synthase III	COG0332 3-oxoacyl-[acyl-carrier-protein] synthase III
IYC_17363	79,659	79,883	225	acyl carrier protein	COG0236 Acyl carrier protein
IYC_17368	80,010	80,936	927	2-nitropropane dioxygenase family oxidoreductase	COG2070 Dioxygenases related to 2-nitropropane dioxygenase
IYC_17373	80,957	81,901	945	malonyl CoA-acyl carrier protein transacylase	COG0331 (acyl-carrier-protein) S-malonyltransferase
IYC_17378	81,919	82,665	747	3-ketoacyl-(acyl-carrier-protein) reductase	COG1028 Dehydrogenases with different specificities(related to short-chain alcohol dehydrogenases)
IYC_17383	82,685	83,923	1,239	3-oxoacyl-(acyl carrier protein) synthase II	COG0304 3-oxoacyl-(acyl-carrier-protein) synthase
IYC_17388	83,927	84,403	477	acetyl-CoA carboxylase, biotin carboxyl carrier protein	COG0511 Biotin carboxyl carrier protein
IYC_17393	84,423	84,857	435	(3R)-hydroxymyristoyl-ACP dehydratase	COG0764 3-hydroxymyristoyl/3-hydroxydecanoyl-(acyl carrier protein) dehydratases
IYC_17398	84,888	86,234	1,347	biotin carboxylase	COG0439 Biotin carboxylase
IYC_17403	86,254	87,123	870	acetyl-coenzyme A carboxylase carboxyl transferase subunit beta	COG0777 Acetyl-CoA carboxylase beta subunit
IYC_17408	87,170	88,009	840	acetyl-CoA carboxylase, carboxyl transferase subunit alpha	COG0825 Acetyl-CoA carboxylase alpha subunit
IYC_17413	88,207	89,154	948	D-isomer specific 2-hydroxyacid dehydrogenase	COG0111 Phosphoglycerate dehydrogenase and related dehydrogenases

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_17418	89,183	90,064	882	LysR family transcriptional regulator	COG0583 Transcriptional regulator
IYC_17423	90,106	91,302	1,197	subtilase family protein	COG1404 Subtilisin-like serine proteases
IYC_17428	91,390	92,295	906	LysR family transcriptional regulator	COG0583 Transcriptional regulator
IYC_17433	92,476	92,907	432	hypothetical protein	COG3238 Uncharacterized protein conserved in bacteria
IYC_17438	93,015	93,500	486	hypothetical protein	COG4508 Uncharacterized protein conserved in bacteria
IYC_17443	93,797	94,474	678	PHP domain-containing protein	COG0613 Predicted metal-dependent phosphoesterases (PHP family)
IYC_17448	94,643	95,887	1,245	group 1 glycosyl transferase family protein	COG0438 Glycosyltransferase
IYC_17453	96,012	96,113	102	hypothetical protein	
IYC_17458	96,187	97,560	1,374	sodium:solute symporter family protein	COG0591 Na+/proline symporter
IYC_17463	97,936	99,129	1,194	GTP-binding protein	COG1160 Predicted GTPases
IYC_17468	99,138	100,514	1,377	thiamine biosynthesis protein ThiH	COG1060 Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes
IYC_17473	100,515	101,573	1,059	biotin synthase	COG0502 Biotin synthase and related enzymes
IYC_17478	101,576	101,827	252	hypothetical protein	
IYC_17483	102,165	103,001	837	radical SAM domain-containing protein	COG1180 Pyruvate-formate lyase-activating enzyme
IYC_17488	102,994	104,490	1,497	hypothetical protein	
IYC_17493	105,047	106,333	1,287	respiratory-chain NADH dehydrogenase family protein	COG4656 Predicted NADH:ubiquinone oxidoreductase, subunit RnfC
IYC_17498	106,436	108,181	1,746	glycine/sarcosine/betaine reductase, component B, subunits alpha and beta	
IYC_17503	108,184	108,375	192	hypothetical protein	
IYC_17508	108,402	109,127	726	D-proline reductase, PrdB subunit	
IYC_17513	109,307	109,777	471	D-proline reductase proprotein PrdA	
IYC_17518	109,835	110,596	762	putative D-proline reductase proprotein PrdA	
IYC_17523	110,893	111,432	540	nitroreductase family protein	COG0778 Nitroreductase
IYC_17528	111,606	112,481	876	methionine aminopeptidase, type I	COG0024 Methionine aminopeptidase
IYC_17533	112,506	112,895	390	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_17538	113,027	113,506	480	rRNA large subunit methyltransferase	COG1576 Uncharacterized conserved protein
IYC_17543	113,687	114,190	504	hypothetical protein	COG3318 Predicted metal-binding protein related to the C-terminal domain of SecA
IYC_17548	114,205	114,786	582	putative lipoprotein	COG5401 Spore germination protein
IYC_17553	114,803	115,588	786	metallo-beta-lactamase family protein	COG1235 Metal-dependent hydrolases of the beta-lactamase superfamily I
IYC_17558	115,614	116,867	1,254	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	COG0766 UDP-N-acetylglucosamine enolpyruvyl transferase
IYC_17563	117,263	117,811	549	hypothetical protein	COG1514 2'-5' RNA ligase
IYC_17568	117,860	119,758	1,899	glutamine synthetase	COG0174 Glutamine synthetase
IYC_17573	119,953	120,726	774	glutamate racemase	COG0796 Glutamate racemase
IYC_17578	120,873	121,391	519	putative peptidyl-prolyl cis-trans isomerase, cyclophilin-type	COG0652 Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family
IYC_17583	121,410	121,784	375	hypothetical protein	
IYC_17588	121,798	122,460	663	DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_17593	122,759	124,555	1,797	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_17598	124,720	125,616	897	radical SAM domain-containing protein	COG1313 Uncharacterized Fe-S protein PflX, homolog of pyruvate formate lyase activating proteins
IYC_17603	125,649	126,695	1,047	hypothetical protein	COG3949 Uncharacterized membrane protein
IYC_17608	126,834	127,952	1,119	putative phage head-tail adaptor	
IYC_17613	128,019	128,951	933	radical SAM family protein	COG1242 Predicted Fe-S oxidoreductase
IYC_17618	128,982	129,701	720	2-phosphosulfolactate phosphatase	COG2045 Phosphosulfolactate phosphohydrolase and related enzymes

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_17623	130,671	131,315	645	haloacid dehalogenase, IA family protein	COG0637 Predicted phosphatase/phosphohexomutase
IYC_17628	131,393	132,742	1,350	[Fe] hydrogenase	COG1145 Ferredoxin
IYC_17633	132,987	133,295	309	hypothetical protein	
IYC_17638	133,360	134,736	1,377	UDP-N-acetylmuramate--L-alanine ligase	COG0773 UDP-N-acetylmuramate-alanine ligase
IYC_17643	134,936	135,751	816	pur operon repressor	COG0503 Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins
IYC_17648	135,854	136,141	288	regulatory protein SpoVG	COG2088 Uncharacterized protein, involved in the regulation of septum location
IYC_17653	136,304	137,677	1,374	bifunctional N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate acetyltransferase	COG1207 N-acetylglucosamine-1-phosphate uridyltransferase (contains nucleotidyltransferase and I-patch acetyltransferase domains)
IYC_17658	137,699	138,658	960	ribose-phosphate pyrophosphokinase	COG0462 Phosphoribosylpyrophosphate synthetase
IYC_17663	138,837	139,523	687	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_17668	139,524	140,933	1,410	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_17673	140,968	142,140	1,173	putative protease	COG0265 Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain
IYC_17678	142,246	142,815	570	peptidyl-tRNA hydrolase	COG0193 Peptidyl-tRNA hydrolase
IYC_17683	142,973	146,479	3,507	transcription-repair coupling factor	COG1197 Transcription-repair coupling factor (superfamily II helicase)
IYC_17688	146,564	147,574	1,011	peptidylprolyl isomerase	COG0760 Parvulin-like peptidyl-prolyl isomerase
IYC_17693	147,772	148,323	552	stage V sporulation protein T	COG2002 Regulators of stationary/sporulation gene expression
IYC_17698	148,456	149,991	1,536	stage V sporulation protein B	COG2244 Membrane protein involved in the export of O-antigen and teichoic acid
IYC_17703	150,006	151,451	1,446	MazG family protein	COG3956 Protein containing tetrapyrrole methyltransferase domain and MazG-like (predicted pyrophosphatase) domain
IYC_17708	151,579	151,857	279	DNA-binding protein HU	COG0776 Bacterial nucleoid DNA-binding protein
IYC_17713	151,933	152,172	240	S4 domain-containing protein	COG1188 Ribosome-associated heat shock protein implicated in the recycling of the 50S subunit (S4 paralog)
IYC_17718	152,290	152,580	291	sporulation protein YabP	
IYC_17723	152,595	152,993	399	spore cortex biosynthesis protein YabQ	
IYC_17728	153,088	153,357	270	putative cell division protein FtsL	
IYC_17733	153,419	153,823	405	hypothetical protein	COG1098 Predicted RNA binding protein (contains ribosomal protein S1 domain)
IYC_17738	154,879	157,266	2,388	stage II sporulation protein E	COG2208 Serine phosphatase RsbU, regulator of sigmasubunit
IYC_17743	157,415	158,812	1,398	tRNA(Ile)-lysidine synthetase	COG0037 Predicted ATPase of the PP-loop superfamilyimplicated in cell cycle control
IYC_17748	158,796	159,341	546	hypoxanthine phosphoribosyltransferase	COG0634 Hypoxanthine-guanine phosphoribosyltransferase
IYC_17753	159,423	161,228	1,806	ATP-dependent metalloprotease FtsH	COG0465 ATP-dependent Zn proteases
IYC_17758	161,432	163,105	1,674	formate--tetrahydrofolate ligase	COG2759 Formyltetrahydrofolate synthetase
IYC_17763	163,260	164,036	777	pantothenate kinase	COG1521 Putative transcriptional regulator, homologof Bvg accessory factor
IYC_17768	164,046	165,017	972	dihydrouridine synthase	COG0042 tRNA-dihydrouridine synthase
IYC_17773	165,067	165,894	828	hypothetical protein	
IYC_17778	166,025	166,507	483	transcription elongation factor GreA	COG0782 Transcription elongation factor
IYC_17783	166,526	168,040	1,515	lysyl-tRNA synthetase	COG1190 Lysyl-tRNA synthetase (class II)
IYC_17788	168,350	169,741	1,392	glycyl-tRNA synthetase	COG0423 Glycyl-tRNA synthetase (class II)
IYC_17793	170,168	170,443	276	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_17798	170,594	171,970	1,377	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	COG0771 UDP-N-acetylmuramoylalanine-D-glutamate ligase
Genomic Scaffold: NZ_JH470556					
IYC_17825	<1	1,007	>1007	hypothetical protein	
IYC_17830	1,027	1,482	456	hypothetical protein	
IYC_17835	2,044	4,821	2,778	enhancing factor	COG3979 Uncharacterized protein contain chitin-binding domain type 3
IYC_17840	5,041	6,624	1,584	hypothetical protein	
IYC_17845	7,335	8,036	702	hypothetical protein	
IYC_17850	8,382	9,467	1,086	Ig group 2 domain-containing protein	COG5492 Bacterial surface proteins containing Ig-like domains
IYC_17855	9,621	10,529	909	hypothetical protein	
IYC_17860	10,629	10,895	267	hypothetical protein	COG0691 tmRNA-binding protein
IYC_17865	11,108	11,293	186	hypothetical protein	
IYC_17870	11,464	12,705	1,242	NLP/P60 protein	COG0791 Cell wall-associated hydrolases (invasion-associated proteins)
IYC_17875	13,228	15,183	1,956	putative tetracycline resistance protein	COG0480 Translation elongation factors (GTPases)
IYC_17880	15,236	16,408	1,173	hypothetical protein	
IYC_17885	16,918	17,277	360	hypothetical protein	
IYC_17890	17,268	17,543	276	hypothetical protein	
IYC_17895	17,892	18,104	213	XRE family transcriptional regulator	COG3655 Predicted transcriptional regulator
IYC_17900	18,117	19,244	1,128	putative DNA methylase	COG0863 DNA modification methylase
IYC_17905	19,247	20,059	813	Modification methylase HpaI	COG0863 DNA modification methylase
IYC_17910	20,189	21,880	1,692	hypothetical protein	
IYC_17915	21,867	22,511	645	hypothetical protein	
IYC_17920	22,794	23,483	690	methionine sulfoxide reductase A	
IYC_17925	23,584	24,087	504	hypothetical protein	COG1637 Predicted nuclease of the RecB family
IYC_17930	24,149	25,369	1,221	transposase, mutator type	COG3328 Transposase and inactivated derivatives
IYC_17935	25,648	26,061	414	hypothetical protein	
IYC_17940	26,262	26,672	411	hypothetical protein	
IYC_17945	26,775	27,149	375	helix-turn-helix domain-containing protein	COG1396 Predicted transcriptional regulators
IYC_17950	27,149	27,934	786	hypothetical protein	
IYC_17955	28,025	28,243	219	helix-turn-helix domain-containing protein	COG1396 Predicted transcriptional regulators
IYC_17960	28,325	28,699	375	hypothetical protein	
IYC_17965	28,699	29,826	1,128	hypothetical protein	
IYC_17970	29,839	30,360	522	hypothetical protein	
IYC_17975	30,405	32,348	1,944	DNA-directed DNA polymerase	COG0749 DNA polymerase I - 3'-5' exonuclease and polymerase domains
IYC_17980	32,404	34,815	2,412	virulence-associated E family protein	COG5545 Predicted P-loop ATPase and inactivated derivatives
IYC_17985	35,090	35,395	306	VRR-NUC domain-containing protein	
IYC_17990	35,392	36,759	1,368	SNF2-related protein	COG0553 Superfamily II DNA/RNA helicases, SNF2 family
IYC_17995	36,752	37,147	396	hypothetical protein	
IYC_18000	37,230	37,442	213	hypothetical protein	
IYC_18005	37,450	37,803	354	prophage LambdaSa04, HNH endonuclease family protein	COG1403 Restriction endonuclease
IYC_18010	38,176	38,616	441	hypothetical protein	
IYC_18015	38,606	39,826	1,221	chromosome partitioning parB family protein	COG1475 Predicted transcriptional regulators
IYC_18020	39,897	40,208	312	hypothetical protein	
IYC_18025	40,385	41,290	906	hypothetical protein	
IYC_18030	41,359	41,568	210	AIG2 family protein	
IYC_18035	41,650	42,132	483	P27 family phage terminase small subunit	COG3747 Phage terminase, small subunit
IYC_18040	42,125	43,672	1,548	Terminase	COG4626 Phage terminase-like protein, large subunit
IYC_18045	43,760	44,206	447	hypothetical protein	
IYC_18050	44,268	45,374	1,107	hypothetical protein	COG1479 Uncharacterized conserved protein
IYC_18055	45,376	45,876	501	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_18060	45,942	47,195	1,254	HK97 family phage portal protein	COG4695 Phage-related protein
IYC_18065	47,195	47,881	687	peptidase s14, clpp	COG0740 Protease subunit of ATP-dependent Clp proteases
IYC_18070	47,895	49,103	1,209	phage major capsid protein HK97 family	COG4653 Predicted phage phi-C31 gp36 major capsid-like protein
IYC_18075	49,127	49,402	276	Bacteriophage QLRG family, putative DNA packaging	
IYC_18080	49,409	49,735	327	phage head-tail adaptor	COG5614 Bacteriophage head-tail adaptor
IYC_18085	49,728	50,126	399	HK97 family phage protein	
IYC_18090	50,119	50,442	324	hypothetical protein	
IYC_18095	50,452	51,021	570	phi13 family phage major tail protein	
IYC_18100	51,114	51,443	330	hypothetical protein	
IYC_18105	51,427	51,852	426	hypothetical protein	
IYC_18110	51,898	52,239	342	hypothetical protein	
IYC_18115	52,287	52,433	147	hypothetical protein	
IYC_18120	52,463	54,982	2,520	TP901 family phage tail tape measure protein	COG5280 Phage-related minor tail protein
IYC_18125	55,572	55,688	117	hypothetical protein	
IYC_18130	55,800	56,714	915	DNA polymerase III subunit epsilon	COG0847 DNA polymerase III, epsilon subunit and related 3'-5' exonucleases
IYC_18135	56,808	58,451	1,644	putative phage-like protein	COG4926 Phage-related protein
IYC_18140	58,507	59,865	1,359	putative phage-like protein	
IYC_18145	59,906	60,313	408	toxin secretion/phage lysis holin	COG4824 Phage-related holin (Lysis protein)
IYC_18150	60,306	61,010	705	N-acetylmuramoyl-L-alanine amidase	COG0860 N-acetylmuramoyl-L-alanine amidase
IYC_18155	61,096	62,385	1,290	recombinase	COG1961 Site-specific recombinases, DNA invertase Pin homologs; overlaps another CDS with the same product name
IYC_18160	62,330	63,661	1,332	recombinase	COG1961 Site-specific recombinases, DNA invertase Pin homologs; overlaps another CDS with the same product name
IYC_18165	63,742	65,463	1,722	hypothetical protein	COG0550 Topoisomerase IA
IYC_18170	65,453	66,427	975	hypothetical protein	
IYC_18175	66,441	67,532	1,092	hypothetical protein	
IYC_18180	67,669	67,860	192	hypothetical protein	
IYC_18185	67,926	68,435	510	RNA polymerase, sigma-24 subunit, ECF subfamily protein	COG0466 ATP-dependent Lon protease, bacterial type
IYC_18190	68,858	72,172	3,315	type I restriction enzyme EcoKI subunit R	COG4096 Type I site-specific restriction-modification system, R (restriction) subunit and related helicases
IYC_18195	72,176	73,636	1,461	N-6 DNA methylase	COG0286 Type I restriction-modification system methyltransferase subunit
IYC_18200	73,636	75,108	1,473	restriction modification system DNA specificity domain-containing protein	COG0732 Restriction endonuclease S subunits
IYC_18205	75,245	77,440	2,196	hypothetical protein	COG0497 ATPase involved in DNA repair
IYC_18210	77,437	78,492	1,056	hypothetical protein	
IYC_18215	78,708	79,736	1,029	hypothetical protein	
IYC_18220	79,888	80,814	927	hypothetical protein	COG2188 Transcriptional regulators
IYC_18225	80,910	81,263	354	PRD domain-containing protein	COG3711 Transcriptional antiterminator
IYC_18230	81,351	81,710	360	hypothetical protein	
IYC_18235	81,741	83,039	1,299	hypothetical protein	
IYC_18240	83,107	83,979	873	aryldialkylphosphatase	COG1735 Predicted metal-dependent hydrolase with the TIM-barrel fold
IYC_18245	84,048	85,157	1,110	hypothetical protein	COG0520 Selenocysteine lyase
IYC_18250	85,203	86,363	1,161	putative amino acid racemase	COG3457 Predicted amino acid racemase
IYC_18255	86,652	87,860	1,209	putative mutase	COG1015 Phosphopentomutase
IYC_18260	88,285	89,010	726	glucosamine-6-phosphate deaminase	COG0363 6-phosphogluconolactonase/Glucosamine-6-phosphate isomerase/deaminase
IYC_18265	89,095	89,253	159	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_18270	89,524	90,141	618	regulatory protein TetR	COG1309 Transcriptional regulator

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_18275	90,227	91,978	1,752	hypothetical protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_18280	91,975	93,714	1,740	hypothetical protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_18285	93,881	94,846	966	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_18290	95,066	95,677	612	hypothetical protein	
IYC_18295	95,674	96,399	726	putative ABC transporter, permease protein	COG0619 ABC-type cobalt transport system, permease component CbiQ and related transporters
IYC_18300	96,413	97,879	1,467	ABC transporter, ATP-binding protein	COG1122 ABC-type cobalt transport system, ATPase component
IYC_18305	97,911	99,701	1,791	ABC transporter	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_18310	99,694	101,436	1,743	ABC transporter	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_18315	101,807	102,421	615	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_18320	102,471	103,823	1,353	Na+driven multidrug efflux pump	COG0534 Na+-driven multidrug efflux pump
IYC_18325	104,035	105,162	1,128	ROK family protein	COG1940 Transcriptional regulator/sugar kinase
IYC_18330	105,374	106,429	1,056	sorbitol dehydrogenase	COG1063 Threonine dehydrogenase and related Zn-dependent dehydrogenases
IYC_18335	106,566	107,960	1,395	sodium:glactoside symporter family protein	COG2211 Na+/melibiose symporter and related transporters
IYC_18340	107,984	109,456	1,473	xylulokinase	COG1070 Sugar (pentulose and hexulose) kinases
IYC_18345	109,870	110,589	720	ABC transporter related protein	COG1122 ABC-type cobalt transport system, ATPase component
IYC_18350	110,595	111,512	918	SufBD protein	COG0719 ABC-type transport system involved in Fe-S cluster assembly, permease component
IYC_18355	111,800	112,924	1,125	radical SAM domain-containing protein	COG2108 Uncharacterized conserved protein related to pyruvate formate-lyase activating enzyme
IYC_18360	113,091	113,222	132	hypothetical protein	
IYC_18365	113,565	114,836	1,272	cytosine permease	COG1457 Purine-cytosine permease and related proteins
IYC_18370	114,880	116,145	1,266	cytosine deaminase	COG0402 Cytosine deaminase and related metal-dependent hydrolases
IYC_18375	116,690	117,967	1,278	lipase	COG1075 Predicted acetyltransferases and hydrolaseswith the alpha/beta hydrolase fold
IYC_18380	118,179	118,748	570	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_18385	119,053	120,846	1,794	myosin-cross-reactive antigen	COG4716 Myosin-crossreactive antigen
IYC_18390	120,973	123,081	2,109	hypothetical protein	COG0714 MoxR-like ATPases
IYC_18395	123,078	124,313	1,236	hypothetical protein	
IYC_18400	124,328	126,271	1,944	hypothetical protein	
IYC_18405	126,809	127,309	501	nitroreductase family protein	COG0778 Nitroreductase
IYC_18410	127,518	127,664	147	hypothetical protein	
IYC_18415	127,675	128,454	780	PhzF family phenazine biosynthesis protein	COG0384 Predicted epimerase, PhzC/PhzF homolog
IYC_18420	128,572	129,966	1,395	GntR family transcriptional regulator	COG1167 Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs
IYC_18425	130,375	131,868	1,494	sodium:neurotransmitter symporter family protein	COG0733 Na+-dependent transporters of the SNF family
IYC_18430	131,885	132,010	126	hypothetical protein	
IYC_18435	132,192	133,073	882	proline iminopeptidase	COG0596 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
IYC_18440	133,154	134,242	1,089	proline dipeptidase	COG0006 Xaa-Pro aminopeptidase
IYC_18445	134,752	136,098	1,347	amino acid/peptide transporter	COG3104 Dipeptide/tripeptide permease
IYC_18450	136,438	137,793	1,356	proton-dependent oligopeptide transporter	COG3104 Dipeptide/tripeptide permease
IYC_18455	138,175	138,570	396	hypothetical protein	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_18460	138,575	138,778	204	DNA-binding protein	COG1476 Predicted transcriptional regulators
IYC_18465	139,049	139,762	714	hypothetical protein	COG1073 Hydrolases of the alpha/beta superfamily
IYC_18470	139,816	139,998	183	hypothetical protein	
IYC_18475	140,089	140,310	222	hypothetical protein	COG4443 Uncharacterized protein conserved in bacteria
IYC_18480	140,589	141,524	936	hypothetical protein	COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily
IYC_18485	141,783	141,995	213	hypothetical protein	COG3478 Predicted nucleic-acid-binding protein containing a Zn-ribbon domain
IYC_18490	142,635	143,300	666	chloramphenicol acetyltransferase	COG4845 Chloramphenicol O-acetyltransferase
IYC_18495	143,571	143,927	357	hypothetical protein	
IYC_18500	144,165	144,821	657	hypothetical protein	COG1309 Transcriptional regulator
IYC_18505	144,899	145,624	726	Polyamine-transporting ATPase	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_18510	145,626	146,345	720	hypothetical protein	
IYC_18515	146,323	146,526	204	hypothetical protein	
IYC_18520	146,549	147,586	1,038	hypothetical protein	COG0477 Permeases of the major facilitator superfamily
IYC_18525	147,802	148,941	1,140	class V aminotransferase	COG0075 Serine-pyruvate aminotransferase/archaeal aspartate aminotransferase
IYC_18530	149,315	150,589	1,275	hypothetical protein	COG3875 Uncharacterized conserved protein
IYC_18535	150,682	151,500	819	polar amino acid ABC transporter amino acid-binding protein	COG0834 ABC-type amino acid transport/signal transduction systems, periplasmic component/domain
IYC_18540	151,570	152,256	687	His/Glu/Gln/Arg/opine ABC transporter permease	COG0765 ABC-type amino acid transport system, permease component
IYC_18545	152,243	152,965	723	polar amino acid ABC transporter ATP-binding protein	COG1126 ABC-type polar amino acid transport system,ATPase component
IYC_18550	152,985	154,136	1,152	aspartate aminotransferase	COG0436 Aspartate/tyrosine/aromatic aminotransferase
IYC_18555	154,643	155,551	909	hypothetical protein	COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily
IYC_18560	155,699	157,045	1,347	GntR family transcriptional regulator	COG1167 Transcriptional regulators containing a DNA-binding HTH domain and an aminotransferase domain (MocR family) and their eukaryotic orthologs
IYC_18565	157,337	158,077	741	SCP-like extracellular protein	COG2340 Uncharacterized protein with SCP/PR1 domains
IYC_18570	158,534	159,658	1,125	DEAD-box ATP dependent DNA helicase	COG0513 Superfamily II DNA and RNA helicases
IYC_18575	160,146	161,921	1,776	putative sigma-54 dependent transcriptional regulator	COG3829 Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding domains
IYC_18580	162,181	163,485	1,305	hypothetical protein	COG3681 Uncharacterized conserved protein
IYC_18585	163,756	164,304	549	GGDEF domain-containing protein	COG2199 FOG: GGDEF domain
IYC_18590	164,902	165,966	1,065	xanthine/uracil permease family protein	COG2252 Permeases
IYC_18595	165,982	166,527	546	adenine phosphoribosyltransferase	COG0503 Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins
IYC_18600	166,927	167,205	279	hypothetical protein	
IYC_18605	167,560	168,621	1,062	alanine racemase, N-terminal domain-containing protein	COG3457 Predicted amino acid racemase
IYC_18610	168,787	170,436	1,650	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_18615	170,607	171,533	927	hypothetical protein	COG1943 Transposase and inactivated derivatives
IYC_18620	171,945	173,036	1,092	glycerol dehydrogenase	COG0371 Glycerol dehydrogenase and related enzymes
IYC_18625	173,475	174,344	870	radical SAM domain-containing protein	COG1032 Fe-S oxidoreductase
IYC_18630	174,545	175,303	759	hypothetical protein	COG0550 Topoisomerase IA
IYC_18635	175,609	176,979	1,371	AMP-dependent synthetase and ligase	COG0147 Anthranilate/para-aminobenzoate synthases component I

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_18640	177,150	178,151	1,002	3-deoxy-7-phosphoheptulonate synthase	COG2876 3-deoxy-D-arabino-heptulosonate 7-phosphate(DAHP) synthase
IYC_18645	178,865	179,272	408	hypothetical protein	
IYC_18650	179,412	180,041	630	putative chloramphenicol acetyltransferase	COG0110 Acetyltransferase (isoleucine patch superfamily)
IYC_18655	180,311	181,048	738	RpiR family transcriptional regulator	COG1737 Transcriptional regulators
IYC_18660	181,249	182,109	861	tagatose-bisphosphate aldolase	COG0191 Fructose/tagatose bisphosphate aldolase
IYC_18665	182,129	183,052	924	PfkB family kinase	COG1105 Fructose-1-phosphate kinase and related fructose-6-phosphate kinase (PfkB)
IYC_18670	183,053	184,447	1,395	PTS system, fructose family, IIBC component	COG1445 Phosphotransferase system fructose-specificcomponent IIB
IYC_18675	184,581	185,021	441	PTS system, fructose family, IIA component	COG1762 Phosphotransferase system mannitol/fructose-specific IIA domain (Ntr-type)
IYC_18680	185,230	185,940	711	trehalose operon repressor	COG2188 Transcriptional regulators
IYC_18685	185,961	187,640	1,680	alpha,alpha-phosphotrehalase	COG0366 Glycosidases
IYC_18690	187,858	189,285	1,428	PTS system, trehalose-specific IIBC component	COG1264 Phosphotransferase system IIB components
IYC_18695	189,762	191,327	1,566	MutS domain-containing protein	COG0249 Mismatch repair ATPase (MutS family)
IYC_18700	191,866	193,746	1,881	heat shock protein 90	COG0326 Molecular chaperone, HSP90 family
IYC_18705	194,273	195,703	1,431	potassium/proton antiporter	COG3263 NhaP-type Na+/H+ and K+/H+ antiporters witha unique C-terminal domain
IYC_18710	195,937	196,692	756	TPR repeat-containing protein	COG0457 FOG: TPR repeat
IYC_18715	196,992	197,972	981	radical SAM domain-containing protein	COG0535 Predicted Fe-S oxidoreductases
IYC_18720	197,985	198,317	333	4-carboxymuconolactone decarboxylase	COG0599 Uncharacterized homolog of gamma-carboxymuconolactone decarboxylase subunit
IYC_18725	198,674	199,102	429	hypothetical protein	
IYC_18730	199,404	199,685	282	hypothetical protein	
IYC_18735	199,873	200,199	327	hypothetical protein	
IYC_18740	200,196	200,561	366	hypothetical protein	
IYC_18745	200,827	201,936	1,110	spore germination protein	COG0531 Amino acid transporters
IYC_18750	201,979	203,091	1,113	Ger(x)C family germination protein	
IYC_18755	203,081	204,169	1,089	spore germination protein	overlaps another CDS with the same product name
IYC_18760	204,162	205,364	1,203	spore germination protein	overlaps another CDS with the same product name
IYC_18765	205,447	206,532	1,086	spore germination protein	COG0531 Amino acid transporters
IYC_18780	208,394	209,461	1,068	alanine racemase domain-containing protein	COG3457 Predicted amino acid racemase
IYC_18785	209,687	210,253	567	putative ABC transporter, permease protein	COG4721 Predicted membrane protein
IYC_18790	210,256	211,689	1,434	ABC transporter, ATP-binding protein	COG1122 ABC-type cobalt transport system, ATPase component
IYC_18795	211,705	212,463	759	putative ABC transporter, permease protein	COG0619 ABC-type cobalt transport system, permease component CbiQ and related transporters
IYC_18800	212,486	213,298	813	hypothetical protein	COG2043 Uncharacterized protein conserved in archaea
IYC_18805	213,808	214,473	666	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_18810	214,463	215,902	1,440	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_18815	216,133	218,352	2,220	hypothetical protein	
IYC_18820	218,373	219,368	996	putative internalin	COG4886 Leucine-rich repeat (LRR) protein
IYC_18825	219,712	221,001	1,290	hypothetical protein	COG3681 Uncharacterized conserved protein
IYC_18830	221,314	222,483	1,170	class I/II aminotransferase	COG1168 Bifunctional PLP-dependent enzyme with beta-cystathionase and maltose regulon repressor activities
IYC_18835	222,529	223,800	1,272	sodium:dicarboxylate symporter family protein	COG1301 Na+/H+-dicarboxylate symporters
IYC_18840	223,845	224,099	255	hypothetical protein	COG4545 Glutaredoxin-related protein

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_18845	224,140	224,508	369	putative endoribonuclease L-PSP	COG0251 Putative translation initiation inhibitor, yjgF family
IYC_18850	224,521	224,775	255	hypothetical protein	COG4545 Glutaredoxin-related protein
IYC_18855	225,028	227,082	2,055	sigma-54 dependent transcriptional regulator	COG3829 Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding domains
IYC_18870	229,497	230,495	999	leucine-rich repeat-containing protein	COG4886 Leucine-rich repeat (LRR) protein
IYC_18875	230,899	231,228	330	hypothetical protein	COG2801 Transposase and inactivated derivatives
IYC_18890	232,343	232,954	612	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_18895	232,951	235,518	2,568	pyruvate phosphate dikinase PEP/pyruvate-bindingprotein	COG0574 Phosphoenolpyruvate synthase/pyruvate phosphate dikinase
IYC_18900	236,461	237,087	627	NADPH-dependent FMN reductase	COG0655 Multimeric flavodoxin WrbA
IYC_18905	237,444	238,265	822	ABC transporter	COG0789 Predicted transcriptional regulators
IYC_18910	238,380	240,110	1,731	ABC transporter, ATP-binding/permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_18915	240,107	241,864	1,758	putative ABC transporter, ATP-binding/permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_18920	242,136	242,951	816	alpha/beta superfamily hydrolase	COG0596 Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)
IYC_18925	243,327	243,485	159	hypothetical protein	
IYC_18930	243,805	245,277	1,473	ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_18935	245,292	245,963	672	ABC transporter, ATP-binding protein	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_18940	246,427	247,071	645	hypothetical protein	COG0398 Uncharacterized conserved protein
IYC_18945	247,235	248,557	1,323	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_18950	248,777	249,628	852	tungstate ABC transporter, tungstate-binding protein	COG2998 ABC-type tungstate transport system, permease component
IYC_18955	249,637	250,308	672	tungstate ABC transporter, permease protein	COG4662 ABC-type tungstate transport system, periplasmic component
IYC_18960	250,305	250,997	693	tungstate ABC transporter ATP-binding protein	COG1126 ABC-type polar amino acid transport system,ATPase component
IYC_18965	251,025	251,513	489	molybdopterin biosynthesis enzyme Mog	COG0521 Molybdopterin biosynthesis enzymes
IYC_18970	251,537	252,763	1,227	molybdopterin biosynthesis enzyme MoeA	COG0303 Molybdopterin biosynthesis enzyme
IYC_18975	252,775	254,679	1,905	putative molybdopterin biosynthesis protein MoeA/LysR substrate binding-domain-containing protein	COG0303 Molybdopterin biosynthesis enzyme
IYC_18980	254,729	255,688	960	molybdenum cofactor biosynthesis protein A	COG2896 Molybdenum cofactor biosynthesis enzyme
IYC_18985	255,692	256,171	480	molybdenum cofactor biosynthesis protein MoaC	COG0315 Molybdenum cofactor biosynthesis enzyme
IYC_18990	256,173	256,613	441	MOSC domain-containing protein	COG2258 Uncharacterized protein conserved in bacteria
IYC_18995	256,663	258,558	1,896	aor transcriptional regulator	COG3284 Transcriptional activator of acetoin/glycerol metabolism
IYC_19000	258,976	260,802	1,827	aldehyde ferredoxin oxidoreductase, tungsten-containing	COG2414 Aldehyde:ferredoxin oxidoreductase
IYC_19005	261,124	261,351	228	molybdopterin converting factor subunit 1	COG1977 Molybdopterin converting factor, small subunit
IYC_19010	261,353	262,036	684	molybdopterin biosynthesis protein MoeB	COG0476 Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2
IYC_19015	262,327	262,767	441	spermine/spermidine acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_19020	262,887	263,330	444	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_19025	263,756	266,083	2,328	putative ABC transporter ATP-binding/permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_19030	266,080	267,942	1,863	ABC transporter ATP-binding/permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_19035	268,534	269,277	744	ABC transporter ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_19040	269,277	270,128	852	ABC transporter, permease protein	COG1277 ABC-type transport system involved in multi-copper enzyme maturation, permease component
IYC_19045	270,289	271,065	777	hypothetical protein	COG1277 ABC-type transport system involved in multi-copper enzyme maturation, permease component
IYC_19050	271,518	272,207	690	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_19055	272,204	273,664	1,461	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_19060	274,129	274,920	792	response regulator	COG0784 FOG: CheY-like receiver
IYC_19065	275,289	275,762	474	methionine sulfoxide reductase A	COG0225 Peptide methionine sulfoxide reductase
IYC_19080	278,277	280,187	1,911	two-component sensor kinase	COG0642 Signal transduction histidine kinase
IYC_19085	280,177	280,878	702	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_19090	280,880	281,530	651	hypothetical protein	
IYC_19095	282,013	283,593	1,581	clostripain	
IYC_19100	283,682	284,671	990	hypothetical protein	
IYC_19105	284,671	285,348	678	putative RNA polymerase sigma factor SigI	COG1191 DNA-directed RNA polymerase specialized sigma subunit
IYC_19110	285,560	286,648	1,089	hypothetical protein	
IYC_19115	<288640	288,888	>249	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_19120	289,391	289,732	342	hypothetical protein	
IYC_19125	290,070	291,473	1,404	glutamate decarboxylase	COG0076 Glutamate decarboxylase and related PLP-dependent proteins
IYC_19130	291,973	292,350	378	HIT family protein	COG0537 Diadenosine tetraphosphate (Ap4A) hydrolaseand other HIT family hydrolases
IYC_19135	292,566	294,071	1,506	hypothetical protein	COG0308 Aminopeptidase N
IYC_19140	294,185	294,925	741	tRNA pseudouridine synthase A	COG0101 Pseudouridylate synthase
IYC_19145	294,996	296,726	1,731	fibronectin-binding protein	COG1293 Predicted RNA-binding protein homologous toeukaryotic snRNP
IYC_19150	296,980	297,801	822	diaminopimelate epimerase	COG0253 Diaminopimelate epimerase
IYC_19155	297,802	298,875	1,074	putative lipoprotein	
IYC_19160	299,204	300,733	1,530	general secretion pathway protein E	COG2804 Type II secretory pathway, ATPase PulE/Tfp pilus assembly pathway, ATPase PilB
IYC_19165	300,870	302,060	1,191	type II secretion system protein F	COG1459 Type II secretory pathway, component PulF
IYC_19170	302,253	302,657	405	putative type IV pilin	
IYC_19175	302,729	303,181	453	hypothetical protein	
IYC_19180	303,162	303,608	447	hypothetical protein	COG2165 Type II secretory pathway, pseudopilin PulG
IYC_19185	303,605	304,045	441	hypothetical protein	COG2165 Type II secretory pathway, pseudopilin PulG
IYC_19190	304,024	304,470	447	putative lipoprotein	
IYC_19195	304,498	305,187	690	hypothetical protein	
IYC_19200	305,203	305,649	447	hypothetical protein	
IYC_19205	305,655	306,122	468	hypothetical protein	
IYC_19210	306,273	306,785	513	shikimate kinase	COG0703 Shikimate kinase
IYC_19215	306,940	307,383	444	3-dehydroquinate dehydratase	COG0757 3-dehydroquinate dehydratase II
IYC_19220	307,472	308,029	558	elongation factor P	COG0231 Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A)
IYC_19225	308,070	308,489	420	hypothetical protein	
IYC_19230	308,686	309,609	924	stage III sporulation protein AA	COG3854 Uncharacterized protein conserved in bacteria
IYC_19235	309,648	310,166	519	stage III sporulation protein SpoAB	
IYC_19240	310,219	310,416	198	stage III sporulation protein AC	
IYC_19245	310,443	310,826	384	stage III sporulation protein AD	
IYC_19250	310,835	311,998	1,164	stage III sporulation protein AE	

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_19255	312,034	312,624	591	stage III sporulation protein AF	
IYC_19260	312,763	313,359	597	stage III sporulation protein AG	
IYC_19265	313,421	313,945	525	stage III sporulation protein AH	
IYC_19270	314,027	314,422	396	asp23 family protein	COG1302 Uncharacterized protein conserved in bacteria
IYC_19275	314,561	314,992	432	transcription antitermination protein NusB	COG0781 Transcription termination factor
IYC_19280	315,063	316,262	1,200	exodeoxyribonuclease VII large subunit	COG1570 Exonuclease VII, large subunit
IYC_19285	316,318	316,533	216	exodeoxyribonuclease VII small subunit	COG1722 Exonuclease VII small subunit
IYC_19290	316,614	317,486	873	geranyltranstransferase	COG0142 Geranylgeranyl pyrophosphate synthase
IYC_19295	318,288	318,857	570	hypothetical protein	
IYC_19300	318,858	320,726	1,869	1-deoxy-D-xylulose-5-phosphate synthase	COG1154 Deoxyxylulose-5-phosphate synthase
IYC_19305	320,748	321,569	822	hemolysin A	COG1189 Predicted rRNA methylase
IYC_19310	321,569	322,408	840	ATP-NAD kinase	COG0061 Predicted sugar kinase
IYC_19315	322,410	322,862	453	arginine repressor	COG1438 Arginine repressor
IYC_19320	322,875	324,578	1,704	DNA repair protein RecN	COG0497 ATPase involved in DNA repair
IYC_19325	324,876	326,156	1,281	O-acetylhomoserine aminocarboxypropyltransferase/cysteine synthase family protein	COG2873 O-acetylhomoserine sulfhydrylase
IYC_19330	326,196	327,422	1,227	homoserine dehydrogenase	COG0460 Homoserine dehydrogenase
IYC_19335	327,449	328,345	897	homoserine O-succinyltransferase	COG1897 Homoserine trans-succinylase
IYC_19340	328,763	329,989	1,227	stage IV sporulation protein B	COG0750 Predicted membrane-associated Zn-dependent proteases 1
IYC_19345	330,412	331,233	822	sporulation transcription factor Spo0A	COG0784 FOG: CheY-like receiver
IYC_19350	331,584	332,006	423	hypothetical protein	
Genomic Scaffold: NZ_JH470557					
IYC_19355	<1	202	>202	spore germination protein	
Genomic Scaffold: NZ_JH470558					
IYC_19360	67	951	885	ferredoxin-NADP(+) reductase subunit alpha	COG0543 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases
IYC_19365	1,102	2,175	1,074	zinc-binding dehydrogenase family oxidoreductase	COG1063 Threonine dehydrogenase and related Zn-dependent dehydrogenases
IYC_19370	2,624	4,540	1,917	carbon monoxide dehydrogenase	COG1151 6Fe-6S prismane cluster-containing protein
IYC_19375	4,957	5,313	357	dinitrogenase iron-molybdenum cofactor family protein	COG1433 Uncharacterized conserved protein
IYC_19380	5,398	6,243	846	CobQ/CobB/MinD/ParA family protein	COG1149 MinD superfamily P-loop ATPase containing an inserted ferredoxin domain
IYC_19385	6,234	7,064	831	iron-sulfur binding protein	COG1149 MinD superfamily P-loop ATPase containing an inserted ferredoxin domain
IYC_19390	7,089	7,448	360	dinitrogenase iron-molybdenum cofactor family protein	COG1433 Uncharacterized conserved protein
IYC_19395	7,897	8,475	579	hypothetical protein	
IYC_19400	8,891	10,009	1,119	VanZ family protein	COG4767 Glycopeptide antibiotics resistance protein
IYC_19405	10,149	11,594	1,446	aminoacyl-histidine dipeptidase	COG2195 Di- and tripeptidases
IYC_19410	12,156	13,184	1,029	putative low specificity L-threonine aldolase	COG2008 Threonine aldolase
IYC_19415	13,177	13,269	93	hypothetical protein	
IYC_19420	13,683	13,817	135	hypothetical protein	
IYC_19425	14,046	14,753	708	hypothetical protein	COG2230 Cyclopropane fatty acid synthase and related methyltransferases
IYC_19430	14,883	15,158	276	hypothetical protein	COG2197 Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain
IYC_19435	15,299	15,577	279	glycerate kinase	COG1929 Glycerate kinase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_19440	15,810	16,607	798	short chain dehydrogenase/reductase family oxidoreductase	COG1028 Dehydrogenases with different specificities(related to short-chain alcohol dehydrogenases)
IYC_19445	16,986	18,164	1,179	putative sodium extrusion protein NatB	COG1668 ABC-type Na+ efflux pump, permease component
IYC_19450	18,178	18,897	720	putative sodium extrusion ATP-binding protein NatA	COG4555 ABC-type Na+ transport system, ATPase component
IYC_19455	19,255	20,151	897	N-acetylmuramoyl-L-alanine amidase	COG0860 N-acetylmuramoyl-L-alanine amidase
IYC_19460	20,434	21,132	699	hypothetical protein	COG2234 Predicted aminopeptidases
IYC_19465	21,378	21,518	141	hypothetical protein	
IYC_19470	21,612	21,746	135	hypothetical protein	
IYC_19475	21,935	22,348	414	hypothetical protein	
IYC_19480	22,547	23,515	969	aliphatic sulfonate ABC transporter substrate-binding protein	COG0715 ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components
IYC_19485	23,668	24,378	711	aliphatic sulfonates ABC transporter ATP-bindingprotein	COG1116 ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component
IYC_19490	24,365	25,135	771	aliphatic sulfonate ABC transporter permease	COG0600 ABC-type nitrate/sulfonate/bicarbonate transport system, permease component
IYC_19495	25,119	25,730	612	hypothetical protein	COG4624 Iron only hydrogenase large subunit, C-terminal domain
IYC_19500	26,005	27,075	1,071	uroporphyrinogen decarboxylase family protein	COG0407 Uroporphyrinogen-III decarboxylase
IYC_19505	27,423	28,871	1,449	FAD/[2Fe-2S]-binding family oxidoreductase	COG0579 Predicted dehydrogenase
IYC_19510	29,223	30,041	819	MerR family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_19515	30,115	31,518	1,404	amino acid permease family protein	COG0531 Amino acid transporters
IYC_19520	31,895	33,643	1,749	xenobiotic ABC transporter ATPase	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_19525	33,645	35,390	1,746	xenobiotic ABC transporter ATPase	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_19530	35,445	36,923	1,479	ABC transporter	COG1122 ABC-type cobalt transport system, ATPase component
IYC_19535	36,925	37,695	771	cobalt transport protein	COG0619 ABC-type cobalt transport system, permease component CbiQ and related transporters
IYC_19540	37,851	38,465	615	hypothetical protein	
IYC_19545	38,586	39,596	1,011	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_19550	39,629	40,363	735	hypothetical protein	COG0500 SAM-dependent methyltransferases
IYC_19555	40,395	41,180	786	pyochelin biosynthetic protein PchC	COG3208 Predicted thioesterase involved in non-ribosomal peptide biosynthesis
IYC_19560	41,177	42,286	1,110	thiazolinyl imide reductase	COG4693 Oxidoreductase (NAD-binding), involved in siderophore biosynthesis
IYC_19565	42,315	43,346	1,032	hypothetical protein	
IYC_19570	43,343	48,307	4,965	pyochelin synthetase F	COG1020 Non-ribosomal peptide synthetase modules and related proteins
IYC_19575	48,304	55,788	7,485	nonribosomal peptide synthetase	COG1020 Non-ribosomal peptide synthetase modules and related proteins
IYC_19580	55,862	56,890	1,029	hypothetical protein	COG1073 Hydrolases of the alpha/beta superfamily
IYC_19585	56,920	58,527	1,608	hypothetical protein	COG1021 Peptide arylation enzymes
IYC_19590	58,578	59,282	705	hypothetical protein	COG2091 Phosphopantetheinyl transferase
IYC_19595	59,287	60,024	738	hypothetical protein	COG3208 Predicted thioesterase involved in non-ribosomal peptide biosynthesis
IYC_19600	60,211	60,753	543	RNase III inhibitor	COG2110 Predicted phosphatase homologous to the C-terminal domain of histone macroH2A1
IYC_19605	61,028	61,291	264	hypothetical protein	COG3326 Predicted membrane protein
IYC_19620	62,117	62,608	492	hypothetical protein	
IYC_19625	62,936	63,766	831	hydroxyethylthiazole kinase	COG2145 Hydroxyethylthiazole kinase, sugar kinase family
IYC_19630	63,711	64,280	570	thiW protein	COG4732 Predicted membrane protein

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_19635	64,723	65,160	438	acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_19640	65,298	66,026	729	metallo-beta-lactamase family protein	COG0491 Zn-dependent hydrolases, including glyoxylases
IYC_19645	66,211	66,483	273	hypothetical protein	COG2197 Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain
IYC_19650	66,767	67,450	684	dithiobiotin synthetase	COG0132 Dethiobiotin synthetase
IYC_19655	67,453	68,409	957	biotin synthase	COG0502 Biotin synthase and related enzymes
IYC_19660	68,402	68,962	561	BioY family protein	COG1268 Uncharacterized conserved protein
IYC_19665	69,101	69,430	330	hypothetical protein	
IYC_19670	69,565	70,161	597	hypothetical protein	
IYC_19675	70,339	72,132	1,794	M24 family metallopeptidase	COG0006 Xaa-Pro aminopeptidase
IYC_19690	73,648	73,896	249	hypothetical protein	
IYC_19695	74,103	74,249	147	hypothetical protein	
IYC_19700	74,347	74,616	270	hypothetical protein	
IYC_19705	74,911	75,894	984	hypothetical protein	COG0543 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases
IYC_19710	75,987	76,154	168	hypothetical protein	
IYC_19715	76,297	76,542	246	hypothetical protein	
IYC_19720	76,864	77,286	423	hypothetical protein	
IYC_19725	77,679	78,272	594	putative lipoprotein	
IYC_19730	78,466	78,915	450	methionine sulfoxide reductase B	COG0229 Conserved domain frequently associated withpeptide methionine sulfoxide reductase
IYC_19735	79,226	79,822	597	superoxide dismutase	COG0605 Superoxide dismutase
IYC_19740	79,957	80,421	465	putative ABC transporter ATP-binding/permease protein	COG1132 ABC-type multidrug transport system, ATPaseand permease components
IYC_19745	80,414	80,821	408	AraC family transcriptional regulator	
IYC_19750	81,249	81,986	738	hypothetical protein	COG1277 ABC-type transport system involved in multi-copper enzyme maturation, permease component
IYC_19755	82,022	82,768	747	hypothetical protein	COG1277 ABC-type transport system involved in multi-copper enzyme maturation, permease component
IYC_19760	82,761	83,681	921	antibiotic ABC transporter, ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_19765	83,929	84,840	912	sensor histidine kinase	COG0642 Signal transduction histidine kinase
IYC_19770	84,837	85,544	708	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_19775	85,841	86,878	1,038	Peptidoglycan-binding lysin domain protein	COG1388 FOG: LysM repeat
IYC_19780	87,176	87,934	759	exonuclease family protein	COG5018 Inhibitor of the KinA pathway to sporulation, predicted exonuclease
IYC_19785	88,229	89,371	1,143	hypothetical protein	COG1852 Uncharacterized conserved protein
IYC_19790	89,511	89,654	144	hypothetical protein	COG1733 Predicted transcriptional regulators
IYC_19795	90,310	90,612	303	stress responsive A/B barrel domain-containing protein	
IYC_19800	90,736	91,323	588	metallo-beta-lactamase family protein	COG1237 Metal-dependent hydrolases of the beta-lactamase superfamily II
IYC_19805	91,607	92,260	654	pentapeptide repeat-containing protein	COG1357 Uncharacterized low-complexity proteins
IYC_19810	92,377	92,775	399	hypothetical protein	COG3910 Predicted ATPase
IYC_19815	93,199	93,504	306	hypothetical protein	
IYC_19820	93,497	94,021	525	hypothetical protein	COG2314 Predicted membrane protein
IYC_19825	94,002	94,454	453	hypothetical protein	
IYC_19830	94,451	94,963	513	ECF subfamily RNA polymerase sigma factor	COG1595 DNA-directed RNA polymerase specialized sigma subunit, sigma24 homolog
IYC_19835	95,597	96,010	414	endoribonuclease L-PSP	COG0251 Putative translation initiation inhibitor, yjgF family
IYC_19840	96,195	96,464	270	hypothetical protein	COG2197 Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain
IYC_19845	96,526	96,642	117	hypothetical protein	
IYC_19850	96,995	97,525	531	hypothetical protein	COG0558 Phosphatidylglycerophosphate synthase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_19855	97,613	98,314	702	pentapeptide repeat family protein	COG1357 Uncharacterized low-complexity proteins
IYC_19860	98,394	99,746	1,353	polyA polymerase family protein	COG0617 tRNA nucleotidyltransferase/poly(A) polymerase
IYC_19865	100,000	100,095	96	hypothetical protein	
IYC_19870	100,468	101,292	825	putative membrane-associated protease	COG1266 Predicted metal-dependent membrane protease
IYC_19875	101,782	102,114	333	PadR family transcriptional regulator	COG1695 Predicted transcriptional regulators
IYC_19880	102,116	102,892	777	membrane protein	COG0772 Bacterial cell division membrane protein
IYC_19885	103,603	103,761	159	hypothetical protein	
IYC_19890	103,911	104,351	441	hypothetical protein	
IYC_19895	104,397	104,864	468	transcription activator effector binding protein	COG3708 Uncharacterized protein conserved in bacteria
IYC_19900	104,891	105,076	186	hypothetical protein	
IYC_19905	105,274	105,666	393	hypothetical protein	COG0295 Cytidine deaminase
IYC_19910	105,704	105,889	186	hypothetical protein	
IYC_19915	106,347	106,529	183	hypothetical protein	
IYC_19920	106,584	106,802	219	hypothetical protein	
IYC_19925	106,808	107,338	531	hypothetical protein	
IYC_19930	107,654	108,100	447	hypothetical protein	COG2606 Uncharacterized conserved protein
IYC_19935	108,280	109,272	993	tryptophanyl-tRNA synthetase II	COG0180 Tryptophanyl-tRNA synthetase
IYC_19940	109,890	110,249	360	hypothetical protein	COG4728 Uncharacterized protein conserved in bacteria
IYC_19945	110,446	110,862	417	hypothetical protein	COG1683 Uncharacterized conserved protein
IYC_19950	110,920	111,012	93	hypothetical protein	
IYC_19955	111,199	112,590	1,392	hypothetical protein	
IYC_19960	113,005	113,583	579	hypothetical protein	
IYC_19965	114,599	115,450	852	hypothetical protein	
IYC_19970	116,022	116,192	171	hypothetical protein	
IYC_19975	116,385	116,576	192	hypothetical protein	
IYC_19980	116,971	117,324	354	hypothetical protein	
IYC_19985	117,338	>118305	>968	hypothetical protein	
Genomic Scaffold: NZ_JH470560					
IYC_19990	937	2,565	1,629	peptidase, M20/M25/M40 family protein	COG4187 Arginine degradation protein (predicted deacylase)
IYC_19995	2,565	3,062	498	hypothetical protein	COG1956 GAF domain-containing protein
IYC_20000	3,401	4,726	1,326	HlyC/CorC family protein	COG1253 Hemolysins and related proteins containing CBS domains
IYC_20005	5,076	5,366	291	hypothetical protein	COG0011 Uncharacterized conserved protein
IYC_20010	5,385	6,698	1,314	thiamine biosynthesis protein ThiC	COG0422 Thiamine biosynthesis protein ThiC
IYC_20015	6,917	8,293	1,377	amidophosphoribosyltransferase	COG0034 Glutamine phosphoribosylpyrophosphate amidotransferase
IYC_20020	8,557	8,838	282	hypothetical protein	COG1550 Uncharacterized protein conserved in bacteria
IYC_20025	8,894	9,583	690	hypothetical protein	
IYC_20030	9,585	10,235	651	hypothetical protein	
IYC_20035	10,242	11,099	858	ABC transporter, ATP-binding protein	COG1131 ABC-type multidrug transport system, ATPasecomponent
IYC_20040	11,314	11,685	372	GntR family transcriptional regulator	COG1725 Predicted transcriptional regulators
IYC_20045	12,393	12,905	513	hypothetical protein	
IYC_20050	13,304	13,954	651	3,4-dihydroxy-2-butanone 4-phosphate synthase	COG0108 3,4-dihydroxy-2-butanone 4-phosphate synthase
IYC_20055	14,858	15,601	744	iron-sulfur cluster-binding protein	COG1145 Ferredoxin
IYC_20060	15,703	16,380	678	cyclic nucleotide-binding protein	COG0664 cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases
IYC_20065	16,595	17,038	444	putative cell wall hydrolase	COG3773 Cell wall hydrolyses involved in spore germination
IYC_20070	17,061	17,570	510	hypothetical protein	
IYC_20075	17,834	18,178	345	hypothetical protein	COG1917 Uncharacterized conserved protein, containsdouble-stranded beta-helix domain

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_20080	18,299	20,461	2,163	putative phage infection protein	COG1511 Predicted membrane protein
IYC_20085	20,784	21,521	738	glycerophosphoryl diester phosphodiesterase family protein	COG0584 Glycerophosphoryl diester phosphodiesterase
IYC_20090	21,761	22,717	957	TPR domain-containing protein	
IYC_20095	23,171	24,424	1,254	decarboxylase family protein	COG0076 Glutamate decarboxylase and related PLP-dependent proteins
IYC_20100	25,147	25,881	735	hypothetical protein	COG1579 Zn-ribbon protein, possibly nucleic acid-binding
IYC_20105	26,020	26,823	804	hypothetical protein	COG0327 Uncharacterized conserved protein
IYC_20110	26,814	27,503	690	hypothetical protein	COG2384 Predicted SAM-dependent methyltransferase
IYC_20115	27,579	28,466	888	hypothetical protein	
IYC_20120	28,541	29,626	1,086	RNA polymerase sigma factor RpoD	COG0568 DNA-directed RNA polymerase, sigma subunit (sigma70/sigma32)
IYC_20125	29,651	31,423	1,773	DNA primase	COG0358 DNA primase (bacterial type)
IYC_20130	31,706	32,734	1,029	deoxyguanosinetriphosphate triphosphohydrolase-like protein	COG0232 dGTP triphosphohydrolase
IYC_20135	32,877	33,947	1,071	spore coat protein-like protein	COG2334 Putative homoserine kinase type II (proteinkinase fold)
IYC_20140	33,994	36,627	2,634	pyruvate phosphate dikinase	COG0574 Phosphoenolpyruvate synthase/pyruvate phosphate dikinase
IYC_20145	36,709	37,350	642	CBS domain-containing protein	COG0517 FOG: CBS domain
IYC_20150	37,853	38,473	621	UBA/TS-N domain-containing protein	COG0264 Translation elongation factor Ts
IYC_20155	38,476	38,853	378	DNA repair protein RecO	COG1381 Recombinational DNA repair protein (RecF pathway)
IYC_20160	39,241	40,131	891	GTPase Era	COG1159 GTPase
IYC_20165	40,147	40,545	399	cytidine deaminase	COG0295 Cytidine deaminase
IYC_20170	40,629	41,327	699	diacylglycerol kinase/PAP2 family protein	COG0818 Diacylglycerol kinase
IYC_20175	41,368	41,868	501	metal-binding heat shock protein	COG0319 Predicted metal-dependent hydrolase
IYC_20180	41,865	43,952	2,088	putative hydrolase	COG1480 Predicted membrane-associated HD superfamily hydrolase
IYC_20185	43,975	45,135	1,161	putative stage IV sporulation protein	
IYC_20190	45,125	45,421	297	sporulation protein YqfC	
IYC_20195	45,537	45,983	447	GatB/Yqey domain-containing protein	COG1610 Uncharacterized conserved protein
IYC_20200	46,026	46,202	177	30S ribosomal protein S21	COG0828 Ribosomal protein S21
IYC_20205	46,339	46,683	345	HIT family protein	COG0537 Diadenosine tetraphosphate (Ap4A) hydrolaseand other HIT family hydrolases
IYC_20210	46,740	48,038	1,299	RNA modification protein	COG0621 2-methylthioadenine synthetase
IYC_20215	48,038	48,796	759	16S ribosomal RNA methyltransferase RsmE	COG1385 Uncharacterized protein conserved in bacteria
IYC_20220	48,893	49,831	939	ribosomal protein L11 methyltransferase	COG2264 Ribosomal protein L11 methylase
IYC_20225	50,007	51,152	1,146	chaperone protein DnaJ	COG0484 DnaJ-class molecular chaperone with C-terminal Zn finger domain
IYC_20230	51,295	53,166	1,872	molecular chaperone DnaK	COG0443 Molecular chaperone
IYC_20235	53,221	53,856	636	heat shock protein GrpE	COG0576 Molecular chaperone GrpE (heat shock protein)
IYC_20240	53,883	54,914	1,032	heat-inducible transcription repressor	COG1420 Transcriptional regulator of heat shock gene
IYC_20245	55,115	56,257	1,143	coproporphyrinogen III oxidase	COG0635 Coproporphyrinogen III oxidase and related Fe-S oxidoreductases
IYC_20250	56,284	58,092	1,809	GTP-binding protein LepA	COG0481 Membrane GTPase LepA
IYC_20255	58,233	58,628	396	hypothetical protein	
IYC_20260	58,699	59,790	1,092	stage II sporulation protein	
IYC_20265	59,921	60,895	975	germination protease	COG0443 Molecular chaperone
IYC_20270	61,105	61,371	267	30S ribosomal protein S20	COG0268 Ribosomal protein S20
IYC_20275	61,406	62,437	1,032	DNA polymerase III subunit delta	COG1466 DNA polymerase III, delta subunit

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_20280	62,466	64,253	1,788	ComEC/Rec2 family protein	COG0658 Predicted membrane metal-binding protein
IYC_20285	64,265	64,495	231	hypothetical protein	
IYC_20290	64,600	67,194	2,595	putative cation transporting ATPase	COG0474 Cation transport ATPase
IYC_20295	67,497	68,894	1,398	alginate O-acetyltransferase	
IYC_20300	68,910	70,325	1,416	alginate O-acetylation protein algI	COG1696 Predicted membrane protein involved in D-alanine export
IYC_20305	72,422	74,329	1,908	selenocysteine-specific translation elongation factor	COG3276 Selenocysteine-specific translation elongation factor
IYC_20310	74,348	75,736	1,389	selenocysteine synthase	COG1921 Selenocysteine synthase [seryl-tRNASer selenium transferase]
IYC_20315	75,802	76,734	933	selenide, water dikinase, selenocysteine-containing	COG0709 Selenophosphate synthase
IYC_20320	76,789	76,881	93	selenide, water dikinase, selenocysteine-containing	
IYC_20325	76,987	77,595	609	putative comE operon protein 1	COG1555 DNA uptake protein and related DNA-binding proteins
IYC_20330	77,784	79,040	1,257	D-alanyl-D-alanine carboxypeptidase family protein	COG1686 D-alanyl-D-alanine carboxypeptidase
IYC_20335	79,156	80,055	900	RluA family pseudouridine synthase	COG0564 Pseudouridylate synthases, 23S RNA-specific
IYC_20340	80,300	80,680	381	putative endoribonuclease L-PSP	COG0251 Putative translation initiation inhibitor, yjgF family
IYC_20345	80,696	81,970	1,275	putative cell envelope-related function transcriptional attenuator	COG1316 Transcriptional regulator
IYC_20350	81,981	82,550	570	HDIG domain-containing protein	COG1713 Predicted HD superfamily hydrolase involvedin NAD metabolism
IYC_20355	82,633	83,238	606	nicotinic acid mononucleotide adenylyltransferase	COG1057 Nicotinic acid mononucleotide adenylyltransferase
IYC_20360	83,429	83,722	294	putative RNA-binding protein	COG1534 Predicted RNA-binding protein containing KHdomain, possibly ribosomal protein
IYC_20365	83,749	85,023	1,275	GTPase CgtA	COG0536 Predicted GTPase
IYC_20370	85,405	85,707	303	50S ribosomal protein L27	COG0211 Ribosomal protein L27
IYC_20375	85,710	86,036	327	hypothetical protein	COG2868 Predicted ribosomal protein
IYC_20380	86,039	86,353	315	50S ribosomal protein L21	COG0261 Ribosomal protein L21
IYC_20385	86,477	87,916	1,440	ribonuclease	COG1530 Ribonucleases G and E
IYC_20390	87,955	88,668	714	hypothetical protein	COG5011 Uncharacterized protein conserved in bacteria
IYC_20395	88,646	90,496	1,851	radical SAM domain-containing protein	COG1032 Fe-S oxidoreductase
IYC_20400	90,496	90,588	93	hypothetical protein	
IYC_20405	90,609	91,460	852	putative metalloprotease	COG1994 Zn-dependent proteases
IYC_20410	91,515	92,300	786	peptidase, family M23/M37	COG0739 Membrane proteins related to metalloendopeptidases
IYC_20415	92,594	93,712	1,119	rod shape-determining protein RodA	COG0772 Bacterial cell division membrane protein
IYC_20420	93,811	94,074	264	cell division topological specificity factor MinE	COG0851 Septum formation topological specificity factor
IYC_20425	94,088	94,885	798	septum site-determining protein MinD	COG2894 Septum formation inhibitor-activating ATPase
IYC_20430	94,903	95,532	630	septum formation inhibitor	COG0850 Septum formation inhibitor
IYC_20435	95,707	98,481	2,775	penicillin-binding protein	COG0768 Cell division protein FtsI/penicillin-binding protein 2
IYC_20440	98,501	98,992	492	rod shape-determining protein MreD	
IYC_20445	99,003	99,848	846	rod shape-determining protein MreC	COG1792 Cell shape-determining protein
IYC_20450	99,855	100,865	1,011	rod shape-determining protein MreB	COG1077 Actin-like ATPase involved in cell morphogenesis
IYC_20455	100,894	101,580	687	hypothetical protein	COG2003 DNA repair proteins
IYC_20460	101,595	102,179	585	Maf-like protein	COG0424 Nucleotide-binding protein implicated in inhibition of septum formation
IYC_20465	102,195	102,452	258	hypothetical protein	
IYC_20470	102,562	103,263	702	hypothetical protein	COG1049 Aconitase B
IYC_20475	103,486	104,025	540	hypothetical protein	
IYC_20480	104,341	104,895	555	hypothetical protein	COG0384 Predicted epimerase, PhzC/PhzF homolog
IYC_20485	104,954	105,391	438	hypothetical protein	COG4492 ACT domain-containing protein
IYC_20490	106,091	106,636	546	transglycosylase SLT domain-containing protein	COG0741 Soluble lytic murein transglycosylase and related regulatory proteins (some contain LysM/invasin domains)
IYC_20495	106,654	107,292	639	dephospho-CoA kinase	COG0237 Dephospho-CoA kinase
IYC_20500	107,304	109,931	2,628	DNA polymerase I	COG0258 5'-3' exonuclease (including N-terminal domain of PolI)
IYC_20505	109,949	111,073	1,125	chorismate synthase	COG0082 Chorismate synthase
IYC_20510	111,559	111,711	153	hypothetical protein	
IYC_20515	111,720	114,104	2,385	hypothetical protein	
IYC_20520	114,115	114,453	339	hypothetical protein	
IYC_20525	114,465	114,890	426	hypothetical protein	
IYC_20530	115,308	115,466	159	hypothetical protein	
IYC_20535	115,657	115,854	198	putative DNA-binding protein	COG1476 Predicted transcriptional regulators
IYC_20540	115,896	116,498	603	hypothetical protein	
IYC_20545	117,914	118,408	495	hypothetical protein	
IYC_20550	118,421	119,110	690	hypothetical protein	
IYC_20555	119,357	119,704	348	N-acetylmuramoyl-L-alanine amidase	COG3103 SH3 domain protein; overlaps another CDS with the same product name
IYC_20560	119,679	120,098	420	N-acetylmuramoyl-L-alanine amidase	COG0860 N-acetylmuramoyl-L-alanine amidase; overlaps another CDS with the same product name
IYC_20565	120,143	>120327	>185	hypothetical protein	
Genomic Scaffold: NZ_JH470561					
IYC_20576	522	944	423	hypothetical protein	
IYC_20581	960	1,481	522	transcriptional regulator, biotin repressor family protein	COG1827 Predicted small molecule binding protein (contains 3H domain)
IYC_20586	1,483	2,646	1,164	iron-sulfur cluster-binding protein	COG2006 Uncharacterized conserved protein
IYC_20591	2,636	5,332	2,697	hypothetical protein	COG1615 Uncharacterized conserved protein
IYC_20596	5,465	6,241	777	ErfK/YbiS/YcfS/YnhG family protein	COG1376 Uncharacterized protein conserved in bacteria
IYC_20601	6,266	7,663	1,398	FAD-dependent family oxidoreductase	COG0579 Predicted dehydrogenase
IYC_20606	7,680	8,033	354	hypothetical protein	COG3862 Uncharacterized protein with conserved CXXCpairs
IYC_20611	8,405	9,685	1,281	seryl-tRNA synthetase	COG0172 Seryl-tRNA synthetase
IYC_20616	10,421	11,584	1,164	phage integrase family site specific recombinase	COG0582 Integrase
IYC_20621	11,585	12,061	477	hypothetical protein	COG1396 Predicted transcriptional regulators
IYC_20626	12,255	12,443	189	hypothetical protein	COG1396 Predicted transcriptional regulators
IYC_20631	12,545	12,718	174	hypothetical protein	
IYC_20636	12,780	13,541	762	hypothetical protein	
IYC_20641	13,551	15,380	1,830	TOPRIM domain protein	COG0358 DNA primase (bacterial type)
IYC_20646	15,405	15,692	288	excisionase family DNA-binding protein	COG0465 ATP-dependent Zn proteases
IYC_20651	15,706	15,945	240	hypothetical protein	COG3617 Prophage antirepressor
IYC_20656	16,270	16,482	213	hypothetical protein	
IYC_20661	16,463	16,672	210	hypothetical protein	
IYC_20666	16,788	16,895	108	hypothetical protein	
IYC_20671	16,908	17,078	171	hypothetical protein	
IYC_20676	17,343	17,753	411	hypothetical protein	
IYC_20681	17,746	19,404	1,659	terminase	COG4626 Phage terminase-like protein, large subunit
IYC_20686	19,417	20,658	1,242	phage portal protein, HK97 family	COG4695 Phage-related protein
IYC_20691	20,642	22,291	1,650	phage major capsid protein (HK97 family)	COG3740 Phage head maturation protease
IYC_20696	22,324	22,644	321	hypothetical protein	
IYC_20701	23,363	23,506	144	hypothetical protein	
IYC_20706	23,563	23,724	162	hypothetical protein	
IYC_20711	23,739	24,152	414	hypothetical protein	COG2801 Transposase and inactivated derivatives
IYC_20716	24,356	25,243	888	phosphatidylserine decarboxylase	COG0688 Phosphatidylserine decarboxylase

locus_tag	Minimum	Maximum	Length	Product	COG Group
IYC_20721	25,346	26,944	1,599	hypothetical protein	COG0661 Predicted unusual protein kinase
IYC_20726	27,021	27,341	321	hypothetical protein	COG3937 Uncharacterized conserved protein
IYC_20731	27,735	29,297	1,563	putative sensory box sigma-54 dependent transcriptional regulator	COG3283 Transcriptional regulator of aromatic aminoacids metabolism
IYC_20736	29,467	30,669	1,203	methionine gamma-lyase	COG0626 Cystathionine beta-lyases/cystathionine gamma-synthases
IYC_20741	30,787	31,233	447	cytidine/deoxycytidylate deaminase family protein	COG0590 Cytosine/adenosine deaminases
IYC_20746	31,869	33,494	1,626	DNA polymerase III subunits gamma and tau	COG2812 DNA polymerase III, gamma/tau subunits
IYC_20751	33,560	33,901	342	hypothetical protein	COG0718 Uncharacterized protein conserved in bacteria
IYC_20756	33,934	34,530	597	recombination protein RecR	COG0353 Recombinational DNA repair protein (RecF pathway)
IYC_20761	34,587	34,850	264	hypothetical protein	
IYC_20766	35,042	35,302	261	sigma-K factor processing regulatory protein BofA	
IYC_20771	35,453	36,145	693	putative DNA-binding transcriptional regulator	COG1489 DNA-binding protein, stimulates sugar fermentation
IYC_20776	36,200	36,685	486	NUDIX family hydrolase	COG2816 NTP pyrophosphohydrolases containing a Zn-finger, probably nucleic-acid-binding
IYC_20781	36,792	37,508	717	hypothetical protein	COG0516 IMP dehydrogenase/GMP reductase
IYC_20786	37,650	38,438	789	iron-sulfur cluster-binding protein	COG1145 Ferredoxin
IYC_20791	38,457	38,618	162	hypothetical protein	
IYC_20796	39,004	39,708	705	thiF domain-containing protein	COG1179 Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 1
IYC_20801	39,792	40,901	1,110	major facilitator superfamily protein	COG0477 Permeases of the major facilitator superfamily
IYC_20806	40,930	41,874	945	GNAT family acetyltransferase	COG1670 Acetyltransferases, including N-acetylases of ribosomal proteins

Appendix C. Core genome of Group I *C. botulinum* and *C. sporogenes* (n=1590 CDS)

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO0001	1	1347	chromosomal replication initiation protein
CBO0002	1610	2713	DNA polymerase III subunit beta
CBO0004	2998	4101	recombination protein F
CBO0005	4122	4385	hypothetical protein
CBO0009	14895	15416	transcriptional regulator
CBO0011	16586	19267	hypothetical protein
CBO0012	19404	20180	ErfK/YbiS/YcfS/YnhG family protein
CBO0013	20205	21602	dehydrogenase
CBO0014	21619	21972	hypothetical protein
CBO0016	22342	23622	seryl-tRNA synthetase
CBO0025	30610	31497	phosphatidylserine decarboxylase
CBO0026	31600	33198	hypothetical protein
CBO0033	40159	40755	recombination protein RecR
CBO0034	40785	41075	hypothetical protein
CBO0035	41225	41485	sigma-K factor processing regulatory protein BofA
CBO0036	41636	42328	sugar fermentation stimulation protein A
CBO0037	42384	42872	hydrolase, NUDIX family
CBO0038	42990	43712	hypothetical protein
CBO0039	43846	44634	(Fe-S)-binding protein
CBO0040	45197	45901	thiF domain-containing protein
CBO0041	45960	47093	major facilitator superfamily protein
CBO0042	47122	48066	GNAT family acetyltransferase
CBO0047	75254	75997	peptidil-prolyl isomerase
CBO0048	81736	82479	NAD-dependent deacetylase
CBO0049	82665	83573	LD-carboxypeptidase
CBO0050	83609	84550	transporter CorA metal ion transporter family
CBO0051	84915	85418	hypothetical protein
CBO0052A	86837	87049	hypothetical protein
CBO0055	89276	89605	hypothetical protein
CBO0057	90569	91483	PSP1 domain-containing protein
CBO0059	91780	91950	ferredoxin, 4Fe-4S
CBO0060	92071	92811	hypothetical protein
CBO0065	97129	98340	arginine deiminase
CBO0066	98426	98893	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
CBO0069	100767	101009	hypothetical protein
CBO0070	101129	101704	spore maturation protein A
CBO0071	101717	102235	spore maturation protein B
CBO0072	102480	103817	MATE efflux family protein
CBO0074	104562	104996	MarR family transcriptional regulator
CBO0075	105378	107315	methionyl-tRNA synthetase
CBO0076	107480	108985	hypothetical protein
CBO0077	109136	109930	TatD family hydrolase
CBO0079	110158	111216	3D/G5 domain protein
CBO0081	111827	112663	dimethyladenosine transferase
CBO0083	114140	114727	hypothetical protein
CBO0085	115897	116958	macrolide ABC transporter
CBO0086	116969	117664	macrolide ABC transporter ATP-binding protein
CBO0089	119933	120124	(Fe-S)-binding protein
CBO0091	120570	120746	hypothetical protein
CBO0093	121951	122679	Mg2+ transporter
CBO0094	122969	125110	anaerobic ribonucleoside triphosphate reductase
CBO0097	126855	127193	PadR family transcriptional regulator
CBO0098	127314	127952	zinc-dependent hydrolase
CBO0099	128388	131510	isoleucyl-tRNA synthetase
CBO0102	137002	138069	mannose-1-phosphate guanylyltransferase
CBO0104	139699	140505	hypothetical protein
CBO0105	140525	141271	hypothetical protein
CBO0106	141496	142878	radical SAM domain protein
CBO0107	143065	144339	sugar-binding protein

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO0110	146747	146941	hypothetical protein
CBO0111	146968	147735	hypothetical protein
CBO0112	147748	148761	spore coat protein
CBO0113	149116	150243	glycosyl transferase family protein
CBO0114	150360	151364	spore coat protein CotS
CBO0115	151488	152390	peptidase
CBO0116	152586	152822	hypothetical protein
CBO0117	152927	154489	LysM domain-containing protein
CBO0119	154869	155669	cyanophycinase
CBO0120	155713	158334	cyanophycin synthetase
CBO0121	158427	159269	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
CBO0123	159692	161212	spore germination protein
CBO0124	161232	162329	spore germination protein
CBO0125	162379	163560	germination protein, Ger(x)C family
CBO0126	163633	164265	stage II sporulation protein R
CBO0127	164302	165543	hypothetical protein
CBO0128	165916	167286	germination protein YpeB
CBO0129	167389	168369	hypothetical protein
CBO0130	168394	168822	hypothetical protein
CBO0131	168954	169253	hypothetical protein
CBO0132	169487	171088	CTP synthetase
CBO0133	171527	171880	hypothetical protein
CBO0135	173537	173755	50S ribosomal protein L31
CBO0136	173982	174557	thymidine kinase
CBO0137	174574	175491	hypothetical protein
CBO0139	176432	177508	peptide chain release factor 1
CBO0140	177536	178144	hypothetical protein
CBO0142	178944	179999	Sua5/YciO/YrdC/YwlC family protein
CBO0144	180520	180957	ribose-5-phosphate isomerase B
CBO0145	180987	181616	uracil phosphoribosyltransferase
CBO0146	182123	182611	cytidine/deoxycytidylate deaminase
CBO0147	182719	183753	UDP-GlcNAc:UDP-P GlcNAc 1-P transferase
CBO0148	183767	184924	UDP-N-acetylglucosamine 2-epimerase
CBO0151	186519	186758	F0F1 ATP synthase subunit C
CBO0153	187316	187855	ATP synthase F0F1 subunit delta
CBO0155	189402	190250	ATP synthase F0F1 subunit gamma
CBO0157	191680	192081	ATP synthase F0F1 subunit epsilon
CBO0161	196139	196885	peptidase
CBO0162	196993	197247	stage III sporulation protein D
CBO0163	197347	198375	rod shape-determining protein Mbl
CBO0168	200776	201951	S-adenosylmethionine synthetase
CBO0172	206499	207026	ribosomal subunit interface protein
CBO0173	207254	209761	preprotein translocase subunit SecA
CBO0175	210125	211102	peptide chain release factor 2
CBO0176	211223	212587	peptidase
CBO0177	212695	214851	RNA-binding protein
CBO0178	215135	215728	hypothetical protein
CBO0180	217800	218069	hypothetical protein
CBO0181	218083	218769	Ser/Thr protein phosphatase
CBO0182	218853	219134	transcription regulator ArsR
CBO0184	221655	222509	hypothetical protein
CBO0185	222666	223454	RelA/SpoT domain-containing protein
CBO0186	223757	226198	leucyl-tRNA synthetase
CBO0189	227842	228441	DnaJ family molecular chaperone
CBO0192	230135	231172	hypothetical protein
CBO0193	231446	232372	fructose-1,6-bisphosphate aldolase
CBO0194	232598	232882	hypothetical protein
CBO0198	235820	236341	CDP-diacylglycerol--serine O-phosphatidyltransferase
CBO0199	236528	236944	hypothetical protein

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO0200	237050	237934	lipid kinase
CBO0201	238008	238652	hypothetical protein
CBO0203	241276	242541	PRC-barrel domain/MgtE domain/CBS domain-containing protein
CBO0204	242771	243466	spore-cortex-lytic enzyme
CBO0206	244759	245355	serine O-acetyltransferase
CBO0208	246490	247191	1-acyl-sn-glycerol-3-phosphate acyltransferase
CBO0209	247284	247925	endonuclease III
CBO0211	249276	250505	competence damage-inducible protein A
CBO0212	250531	251514	hypothetical protein
CBO0213	251624	252658	glutamyl aminopeptidase
CBO0216	255650	256657	spore photoproduct lyase
CBO0217	256785	257252	hypothetical protein
CBO0218	257265	258104	DegV family protein
CBO0219	258398	259501	lipoprotein, bmp family
CBO0221	261219	262322	sugar ABC transporter permease
CBO0222	262315	263241	sugar ABC transporter permease
CBO0224	264407	265780	RNA polymerase factor sigma-54
CBO0225	266068	267099	central glycolytic genes regulator
CBO0226	267177	268184	glyceraldehyde-3-phosphate dehydrogenase, type I
CBO0227	268482	269678	phosphoglycerate kinase
CBO0228	269740	270486	triosephosphate isomerase
CBO0229	270578	272107	phosphoglyceromutase
CBO0230	272163	273458	phosphopyruvate hydratase
CBO0232	274123	276300	ribonuclease R
CBO0233	276709	277467	stationary phase survival protein SurE
CBO0234	277724	278194	SsrA-binding protein
CBO0235	278333	279496	sodium:dicarboxylate symporter family protein
CBO0245	290562	291671	hypothetical protein
CBO0261	309547	310017	hypothetical protein
CBO0264	311324	312001	DNA-binding response regulator
CBO0274	322585	323355	peptide ABC transporter Pep4E family, ATP-binding protein
CBO0275	323369	325339	peptide ABC transporter Pep4E family, permease
CBO0277	327313	328884	PTS system, glucose/glucoside family, IIBC component
CBO0278	329200	330525	maltose-6'-phosphate glucosidase
CBO0279	330839	331348	PTS system glucose/glucoside family transporter subunit IIA
CBO0280	331631	332392	transcriptional regulator
CBO0282	332880	333077	cold shock protein
CBO0285	334991	335596	2-hydroxychromene-2-carboxylate isomerase
CBO0287	336699	337460	transcriptional regulator
CBO0288	337567	339162	PTS system glucose/maltose-specfic transporter subunit IIBC
CBO0290	340771	341244	PTS system glucose/maltose-specific transporter subunit IIA
CBO0291	341382	343097	dihydroxy-acid dehydratase
CBO0292	343149	343796	2-dehydro-3-deoxyphosphogluconate aldolase
CBO0293	343798	344802	2-dehydro-3-deoxygluconokinase
CBO0295	347264	347941	DNA-binding response regulator
CBO0297	349611	350579	glycosyl transferase family protein
CBO0298	350563	350982	cell wall teichoic acid glycosylation protein GtcA
CBO0303	357186	357956	iron ABC transporter ATP-binding protein
CBO0304	357992	359320	drug/sodium antiporter
CBO0317	371460	372032	xanthine phosphoribosyltransferase
CBO0318	372483	373175	DNA-binding response regulator
CBO0319	373188	374168	sensor histidine kinase
CBO0320	374344	375201	UDP pyrophosphate phosphatase
CBO0322	375962	377824	peptide ABC transporter Pep4E family, permease
CBO0326	379671	380321	lipoprotein
CBO0329	381753	383261	hypothetical protein
CBO0330	383585	384136	hypothetical protein
CBO0331	384315	384782	hypothetical protein
CBO0332	384775	385464	accessory gene regulator protein B

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO0332A	385448	385582	lipoprotein
CBO0334	387588	388097	GNAT family acetyltransferase
CBO0336	388973	390820	two-component sensor histidine kinase
CBO0341	394506	395372	PP-loop family protein
CBO0342	395496	396041	DNA-binding protein
CBO0343	396396	397802	amino acid permease
CBO0344	397921	399003	metallopeptidase
CBO0345	399431	402019	bifunctional acetaldehyde-CoA/alcohol dehydrogenase
CBO0346	402338	402625	hypothetical protein
CBO0347	402688	403278	hypothetical protein
CBO0349	404386	405396	carbohydrate-binding protein
CBO0350	405500	406027	PAP2 family protein
CBO0351	406438	407250	polysaccharide deacetylase
CBO0352	407358	408044	DNA-binding response regulator
CBO0353	408054	409448	sensor histidine kinase
CBO0354	409511	411001	lipoprotein
CBO0355	411045	411611	hypothetical protein
CBO0356	411637	412311	DedA family protein
CBO0360	418302	418868	metal dependent phosphohydrolase
CBO0362	420582	421904	class V aminotransferase
CBO0364	423432	424106	hypothetical protein
CBO0367	426362	427777	sodium:solute symporter family protein
CBO0369	430343	431275	RnfABCDGE type electron transport complex subunit D
CBO0370	431275	431946	electron transport complex, RnfABCDGE type, G subunit
CBO0372	432610	433188	RnfABCDGE type electron transport complex subunit A
CBO0373	433202	434035	ferredoxin
CBO0374	434381	435469	amidase domain protein
CBO0375	435663	436199	hypothetical protein
CBO0377	436724	437758	apbE family protein
CBO0399	468193	470247	methyl-accepting chemotaxis protein
CBO0400	470844	472574	cell division protein
CBO0401	472794	473849	hypothetical protein
CBO0402	473824	474273	ribonuclease H
CBO0403	474336	474512	hypothetical protein
CBO0404	474693	475997	tetratricopeptide repeat/DNA binding domain-containing protein
CBO0405	476029	477282	tetratricopeptide repeat protein
CBO0410	482810	483538	CobB/CobQ family glutamine amidotransferase
CBO0412	485222	486202	mannose-6-phosphate isomerase
CBO0413	486551	487666	lipoprotein
CBO0414	487774	488628	UDP pyrophosphate phosphatase
CBO0415	494523	495386	metallo beta-lactamase superfamily lipoprotein
CBO0416	495523	496245	1-acyl-sn-glycerol-3-phosphate acyltransferase
CBO0421	503006	504772	Na+/H+ antiporter family protein
CBO0422	504799	505014	hypothetical protein
CBO0423	505102	505485	aspartate alpha-decarboxylase
CBO0426	508161	511613	ATP-dependent nuclease subunit B
CBO0428	512520	513170	tetratricopeptide repeat protein
CBO0429	513173	513859	hypothetical protein
CBO0436	521001	524840	recombination helicase AddA
CBO0438	526748	527974	peptidase T
CBO0443	530675	531148	hypothetical protein
CBO0446	533826	534629	polar amino acid ABC transporter amino acid-binding protein
CBO0447	534850	535578	hypothetical protein
CBO0448	535927	537138	diguanylate phosphodiesterase
CBO0451	539035	539835	phosphomethylpyrimidine kinase
CBO0452	540847	541749	D-alanine--D-alanine ligase
CBO0453	541905	543047	high-affinity branched-chain amino acid ABC transporter amino acid-binding protein
CBO0454	543051	543923	high-affinity branched-chain amino acid ABC transporter permease
CBO0455	543924	544868	high-affinity branched-chain amino acid ABC transporter permease

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO0456	544880	545617	high-affinity branched chain amino acid ABC transporter ATP-binding protein
CBO0458	546349	547125	hypothetical protein
CBO0460	548279	548722	hypothetical protein
CBO0461	548864	550315	GntR family transcriptional regulator
CBO0462	550376	550945	hypothetical protein
CBO0466	554471	554953	hypothetical protein
CBO0467	555129	555653	hypothetical protein
CBO0472	557819	559048	nuclease SbcCD subunit D
CBO0474	562581	563129	hydrolase, NUDIX family
CBO0475	563940	564791	hypothetical protein
CBO0477	565256	565621	transcription regulator ArsR
CBO0478	565697	567925	cadmium-translocating P-type ATPase
CBO0483	572462	573838	PhoH family protein
CBO0484	574106	575110	exported protein
CBO0486	576447	576629	hypothetical protein
CBO0487	576794	577750	streptolysin associated protein SagB
CBO0491	580831	581559	hypothetical protein
CBO0492	581868	582800	ABC transporter ATP-binding protein
CBO0493	582803	583936	ABC transporter permease/ATP-binding protein
CBO0495	585208	585858	metalloprotease
CBO0496	586326	586796	hypothetical protein
CBO0499	589895	590548	tRNA (guanine-N(7)-)-methyltransferase
CBO0500	590761	591024	hypothetical protein
CBO0501	591119	591613	thiol peroxidase
CBO0502	591984	592727	3-ketoacyl-ACP reductase
CBO0503	592790	593743	response regulator
CBO0505	596331	597275	hydrolase, alpha/beta fold family
CBO0507	599082	599666	HDIG domain protein
CBO0509	601931	602725	response regulator
CBO0510	603207	604484	seryl-tRNA synthetase
CBO0511	605062	607416	ATP-dependent helicase
CBO0512	607707	607898	hypothetical protein
CBO0513	608228	608848	GntR family transcriptional regulator
CBO0515	609103	610056	N(5)-(carboxyethyl)ornithine synthase
CBO0516	610260	612266	fructose-1,6-bisphosphatase
CBO0518	613087	613500	hemerythrin-like metal-binding domain-containingprotein
CBO0520	615148	615570	hypothetical protein
CBO0522	616689	617528	metallo-beta-lactamase
CBO0523	617626	617925	hypothetical protein
CBO0526	621121	621585	transcription regulator AsnC
CBO0527	622110	622484	hypothetical protein
CBO0528	623108	624481	amino acid permease
CBO0530	626620	627129	transferase, hexapeptide repeat family
CBO0531	627174	628046	metallo beta-lactamase superfamily hydrolase
CBO0532	628367	630985	calcium-translocating P-type ATPase
CBO0534	631815	632351	chromate transporter
CBO0535	632525	633115	PemK family protein
CBO0536	633562	633987	hypothetical protein
CBO0537	634193	636109	NAD synthetase
CBO0538	636331	636771	hypothetical protein
CBO0539	636774	637769	hypothetical protein
CBO0542	639161	640396	xaa-pro aminopeptidase
CBO0550	645801	647444	carbon starvation protein CstA
CBO0552	649514	651505	methyl-accepting chemotaxis protein
CBO0555	653982	654383	large conductance mechanosensitive channel protein
CBO0559	657953	658648	ABC transporter ATP-binding protein
CBO0560	658648	659409	ABC transporter permease
CBO0562	661308	662267	spermidine/putrescine ABC transporter permease
CBO0575	673955	674656	DNA-binding response regulator, LytTr family

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO0576	675270	675389	hypothetical protein
CBO0577	675429	675554	hypothetical protein
CBO0578	676248	676394	hypothetical protein
CBO0580	676913	678265	radical SAM domain protein
CBO0581	678258	678848	hypothetical protein
CBO0582	678885	679382	hypothetical protein
CBO0583	679417	681129	ABC transporter permease/ATP-binding protein
CBO0584	681239	681988	lipoprotein
CBO0585	682011	682490	GNAT family acetyltransferase
CBO0587	684970	685650	ABC transporter ATP-binding protein
CBO0588	686198	686482	hypothetical protein
CBO0590	686811	687677	degV family protein
CBO0591	688075	688587	hypothetical protein
CBO0595	692876	694099	nucleoside transporter NupC family
CBO0597	695272	696663	asparaginyl-tRNA synthetase
CBO0609	707816	708532	D-alanyl-D-alanine carboxypeptidase
CBO0610	708553	709245	DNA-binding response regulator
CBO0612	710516	711655	hypothetical protein
CBO0613	711798	712610	response regulator
CBO0622	722626	722955	chaperone CsaA
CBO0624	724185	724736	DNA-binding protein
CBO0625	724993	725583	hypothetical protein
CBO0630	730100	731995	MutS domain-containing protein
CBO0643	738200	739477	ATP-dependent RNA helicase RhlE
CBO0644	740053	740541	rubrerythrin
CBO0647	742733	743923	hypothetical protein
CBO0648	744048	744923	LysR family transcriptional regulator
CBO0654	752190	753029	BglG family transcriptional antiterminator
CBO0655	753234	755075	PTS system beta-glucoside-specific transporter subunit IIABC
CBO0656	755087	755905	PHP domain-containing protein
CBO0657	756532	757242	CRP family transcriptional regulator
CBO0658	757401	758633	metallo-beta-lactamase
CBO0659	758693	759391	scdA protein
CBO0660	759417	760172	hypothetical protein
CBO0661	760245	760664	hypothetical protein
CBO0663	763125	763592	hypothetical protein
CBO0666	765849	766835	metallo-beta-lactamase superfamily protein
CBO0676	775171	775635	transcription elongation factor GreA
CBO0688	790421	791461	ferrichrome ABC transporter permease
CBO0689	791689	792474	ferrichrome ABC transporter ATP-binding protein
CBO0690	792504	793253	nitrogenase iron protein
CBO0693	796125	798137	methyl-accepting chemotaxis protein
CBO0697	801658	802050	glycine cleavage system protein H
CBO0698	802053	803393	glycine dehydrogenase subunit 1
CBO0699	803393	804847	glycine dehydrogenase subunit 2
CBO0701	806447	807424	lipoate-protein ligase A
CBO0716	825962	826222	tfoX domain protein
CBO0719	828168	829634	methyl-accepting chemotaxis protein
CBO0724	833324	834232	cysteine synthase
CBO0725	834216	834668	S-ribosylhomocysteinase
CBO0727	835913	836539	hypothetical protein
CBO0730	840021	840353	hypothetical protein
CBO0731	840441	840941	hypothetical protein
CBO0732	841475	842083	hypothetical protein
CBO0733	842176	842598	DNA-binding protein
CBO0735	843044	844303	phage replication protein
CBO0740	846935	847156	hypothetical protein
CBO0741	847176	847403	hypothetical protein
CBO0753	855490	855813	arsenical resistance operon repressor

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO0767A	868034	868495	hypothetical protein
CBO0768	868485	869459	thymidylate synthase
CBO0769	869471	870229	UbiE/COQ5 family methyltransferase
CBO0770	870275	871489	thiaminase I
CBO0771	871567	872394	phosphomethylpyrimidine kinase
CBO0772	872566	872799	hypothetical protein
CBO0773	873008	873472	hypothetical protein
CBO0781	882469	883464	hypothetical protein
CBO0783	884953	885645	ABC transporter ATP-binding protein
CBO0784	885632	888298	ABC transporter permease
CBO0786	888856	889569	DNA-binding response regulator
CBO0787	889723	890985	sensor histidine kinase
CBO0788	891175	891816	hypothetical protein
CBO0790	892425	893204	hypothetical protein
CBO0791	893870	894667	dihydrodipicolinate reductase
CBO0792	894689	895528	sulfurtransferase
CBO0811	915784	917181	hypothetical protein
CBO0815	922099	924579	methyl-accepting chemotaxis protein/ extracellular solute-binding protein, family 5
CBO0819	928200	928814	alpha-ribazole-5'-phosphate phosphatase
CBO0823	933081	933362	hypothetical protein
CBO0824	933493	933675	hypothetical protein
CBO0825	933693	933941	hypothetical protein
CBO0826	934510	935862	hypothetical protein
CBO0827	936133	936486	6-pyruvoyl tetrahydrobiopterin synthase
CBO0830	938260	938715	2-amino-4-hydroxy-6- hydroxymethyldihydropteridine pyrophosphokinase
CBO0833	941001	942176	exopolysaccharide biosynthesis protein
CBO0851	958982	959635	pentapeptide repeat-containing protein
CBO0863	967797	969239	secreted lipase
CBO0883	988122	988337	hypothetical protein
CBO0891	995788	995967	hypothetical protein
CBO0894	997879	999111	peptidyl-arginine deiminase
CBO0905	1009866	1010537	DNA-binding response regulator
CBO0910	1015542	1017149	LexA repressor
CBO0913	1018406	1019170	MerR family transcriptional regulator
CBO0915	1021359	1022321	cobalamin biosynthesis protein
CBO0916	1022540	1023628	aminotransferase
CBO0917A	1024625	1024924	cobalt ABC transporter permease
CBO0923	1029682	1030656	delta-aminolevulinic acid dehydratase
CBO0924	1030777	1032099	cobyrinic acid a,c-diamide synthase
CBO0931	1037516	1037944	GNAT family acetyltransferase
CBO0933	1038926	1039447	nitroreductase
CBO0936	1040525	1041604	cobalt-precorrin-6A synthase
CBO0939	1043354	1043929	precorrin-6Y C5,15-methyltransferase subunit CbiT
CBO0946	1047486	1048202	cobalt-precorrin-2 C(20)-methyltransferase
CBO0949	1050486	1051217	precorrin-3B C(17)-methyltransferase
CBO0950	1051351	1052118	cobalt-precorrin-6x reductase
CBO0951	1052198	1052686	hypothetical protein
CBO0953	1053657	1054496	kinase
CBO0954	1054644	1055165	cob(I)yrinic acid a,c-diamide adenosyltransferase
CBO0956	1056197	1056961	formate/nitrate transporter
CBO0957	1056981	1058009	anaerobic sulfite reductase subunit A
CBO0959	1058806	1059771	anaerobic sulfite reductase subunit C
CBO0960	1060125	1061147	molybdopterin biosynthesis protein
CBO0966	1065317	1066336	hypothetical protein
CBO0968	1067248	1068024	hypothetical protein
CBO0971	1069381	1070802	sensor histidine kinase
CBO0972	1071180	1071950	iron ABC transporter ATP-binding protein
CBO0974	1073031	1074053	iron chelate uptake ABC transporter solute-binding protein
CBO0975	1074611	1076371	dihydroxyacetone kinase

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO0976	1076382	1077512	glycerol dehydrogenase
CBO0981	1079940	1080383	GNAT family acetyltransferase
CBO0982	1080536	1081129	hypothetical protein
CBO0984	1081513	1082358	oxidoreductase, aldo/keto reductase
CBO0985	1082459	1082782	TfoX domain-containing protein
CBO0986	1082828	1083682	GNAT family acetyltransferase
CBO0992	1088016	1090184	phage infection protein
CBO0993	1090210	1092369	phage infection protein
CBO0996	1096523	1096954	ferritin
CBO0997	1097237	1097782	phosphodiesterase
CBO1013	1114866	1115411	cyclase
CBO1014	1115712	1116896	aspartate aminotransferase
CBO1016	1118562	1119086	MarR family transcriptional regulator
CBO1018	1120425	1121123	B3/4 domain-containing protein
CBO1019	1121388	1122677	aspartyl-tRNA synthetase
CBO1020	1123222	1123602	hypothetical protein
CBO1021	1123762	1124376	TetR family transcriptional regulator
CBO1022	1124392	1124880	hypothetical protein
CBO1023	1124987	1125676	HAD-superfamily hydrolase
CBO1025	1126794	1127378	DJ-1/Pfpl family protein
CBO1026	1127534	1128427	LysR family transcriptional regulator
CBO1030	1130352	1132508	ferrous iron transport protein B
CBO1031	1132529	1132675	hypothetical protein
CBO1033	1133974	1134774	D-alanyl-D-alanine carboxypeptidase
CBO1035	1135518	1136204	DNA-binding response regulator
CBO1036	1136205	1137560	sensor histidine kinase
CBO1038	1138095	1139354	cytochrome P450
CBO1039	1139494	1141302	heavy metal-associated domain-containing protein
CBO1040	1141839	1142075	hypothetical protein
CBO1041	1142345	1143046	DNA-binding response regulator
CBO1045	1146718	1147671	iron chelate uptake ABC transporter FeCT family,permease
CBO1045A	1147661	1148611	iron ABC transporter permease
CBO1046	1148608	1149363	iron ABC transporter ATP-binding protein
CBO1047	1149437	1150381	iron ABC transporter substrate-binding protein
CBO1052	1155823	1157196	sensor histidine kinase
CBO1053	1157199	1157906	DNA-binding response regulator
CBO1054	1157994	1158869	hypothetical protein
CBO1057	1163311	1164810	extracellular exochitinase
CBO1059	1167503	1168498	hypothetical protein
CBO1060	1168663	1169247	hypothetical protein
CBO1061	1169560	1170261	CAAX amino terminal protease
CBO1062	1170377	1171183	metallo-beta-lactamase
CBO1064	1174124	1175338	hypothetical protein
CBO1065	1175713	1176063	ArsC family protein
CBO1068	1177382	1178800	FAD-dependent oxidoreductase
CBO1069	1178857	1180113	pyridine nucleotide-disulfide oxidoreductase
CBO1070	1180116	1180484	hypothetical protein
CBO1071	1180573	1181910	glycerol-3-phosphate transporter
CBO1072	1182492	1184183	arginyl-tRNA synthetase
CBO1073	1184628	1185758	transaminase
CBO1076	1186669	1188432	ferrous iron transport protein B
CBO1077	1188447	1188668	ferrous ion transport protein A
CBO1079	1189602	1190327	haloacid dehalogenase
CBO1085	1194478	1195119	methionine ABC transporter permease
CBO1086	1195174	1195986	lipoprotein, NLPA family
CBO1087	1196175	1196987	lipoprotein, NLPA family
CBO1088	1197215	1198381	amidohydrolase
CBO1091	1200088	1200780	GntR family transcriptional regulator
CBO1094	1202550	1204007	glutamyl-tRNA synthetase

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO1096	1205693	1207201	metallo-beta-lactamase
CBO1098	1207859	1208503	HD domain-containing protein
CBO1103	1214275	1215642	MATE efflux family protein
CBO1105	1216705	1217682	hypothetical protein
CBO1107	1218489	1219727	ABC transporter solute-binding protein
CBO1109	1220618	1221424	ABC transporter permease
CBO1110	1221436	1222455	ABC transporter ATP-binding protein
CBO1111	1222758	1223717	hypothetical protein
CBO1117	1227570	1228553	bifunctional biotin operon repressor/biotin--[acetyl-CoA-carboxylase] synthetase
CBO1118	1228717	1228920	heavy metal-binding protein
CBO1119	1229196	1229633	hypothetical protein
CBO1121	1231338	1231463	hypothetical protein
CBO1122	1232092	1232871	hypothetical protein
CBO1124	1233348	1234706	magnesium transporter
CBO1126	1235709	1236782	class V aminotransferase
CBO1127	1237010	1237954	D-3-phosphoglycerate dehydrogenase
CBO1128	1237932	1239254	hypothetical protein
CBO1129	1239855	1241045	metallo-beta-lactamase/flavodoxin domain-containing protein
CBO1130	1241204	1245493	CoA-substrate-specific enzyme activase
CBO1131	1245567	1247081	choline/carnitine/betaine transporter
CBO1132	1247440	1248066	glycine betaine uptake system transcriptional repressor
CBO1133	1248177	1248677	hypothetical protein
CBO1135	1249694	1250005	phage shock protein
CBO1136	1250081	1250545	hypothetical protein
CBO1137	1250822	1251022	hypothetical protein
CBO1139	1252271	1252843	xanthine phosphoribosyltransferase
CBO1143	1256373	1257476	transporter
CBO1145	1258122	1259861	hypothetical protein
CBO1147	1260954	1262153	hypothetical protein
CBO1154	1267332	1269020	ABC transporter permease
CBO1160	1273992	1274411	thioesterase
CBO1161	1274583	1274999	hypothetical protein
CBO1162	1275868	1276875	ABC transporter substrate-binding protein
CBO1163	1276875	1277780	ABC transporter permease
CBO1164	1277743	1278537	ABC transporter ATP-binding protein
CBO1167	1281371	1282252	hypothetical protein
CBO1168	1282835	1283932	kinase, PfkB family
CBO1169	1283926	1284849	indigoidine synthase A family protein
CBO1170	1285341	1286252	2-dehydropantoate 2-reductase
CBO1173	1288355	1289149	HAD hydrolase
CBO1177	1291175	1291984	zinc transporter
CBO1178	1292656	1293093	rrf2 family protein
CBO1179	1293422	1294351	cysteine synthase A
CBO1180	1294396	1295064	amino acid ABC transporter permease
CBO1183	1297226	1298026	hypothetical protein
CBO1185	1298425	1299495	tRNA-specific 2-thiouridylase MnmA
CBO1192	1306966	1310484	pyruvate-flavodoxin oxidoreductase
CBO1202	1320890	1322335	glycosyl hydrolase family protein
CBO1204	1325403	1326605	tyrosyl-tRNA synthetase
CBO1206	1327256	1328128	ABC transporter ATP-binding protein
CBO1208	1329009	1329377	hypothetical protein
CBO1214	1336966	1337640	L-serine dehydratase, iron-sulfur-dependent subunit beta
CBO1220	1340356	1341324	dipeptidase
CBO1225	1345488	1346345	oligopeptide ABC transporter permease
CBO1226	1346342	1347295	oligopeptide ABC transporter permease
CBO1227	1347494	1349092	oligopeptide ABC transporter oligopeptide-binding protein
CBO1230	1349882	1350454	spore coat peptide assembly protein CotJC
CBO1231	1350684	1351703	hypothetical protein
CBO1233	1353431	1353922	hypothetical protein

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO1238	1356296	1357051	hypothetical protein
CBO1239	1357433	1357591	hypothetical protein
CBO1240	1357709	1358242	DNA-binding protein
CBO1241	1358666	1359715	hypothetical protein
CBO1245	1362004	1363440	sodium:alanine symporter family protein
CBO1247	1364610	1364918	hypothetical protein
CBO1251	1367204	1367950	lipoprotein
CBO1253	1368528	1369988	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase
CBO1255	1371766	1372143	GrdX protein
CBO1256	1372234	1373520	glycine reductase complex component B, alpha andbeta subunits
CBO1260	1375980	1376300	thioredoxin family protein
CBO1263	1376957	1378495	glycine reductase complex component C subunit beta
CBO1264	1378511	1379677	glycine reductase complex component C subunit alpha
CBO1265	1380104	1380715	3-methyladenine DNA glycosylase
CBO1273	1387915	1388274	hypothetical protein
CBO1275	1389751	1390479	hypothetical protein
CBO1276	1390641	1391006	hypothetical protein
CBO1284	1395838	1396143	hypothetical protein
CBO1285	1396200	1398497	xanthine dehydrogenase subunit XdhA
CBO1286	1398511	1399395	xanthine dehydrogenase subunit XdhB
CBO1289	1400879	1403074	DNA topoisomerase III
CBO1291	1404057	1404293	hypothetical protein
CBO1292	1405020	1405673	translaldolase
CBO1298	1413747	1415843	cellulose synthase domain protein
CBO1303	1419160	1419426	hypothetical protein
CBO1305	1422021	1422236	copper chaperone CopZ
CBO1308	1425468	1427186	ABC transporter permease/ATP-binding protein
CBO1311	1429257	1430150	DegV family protein
CBO1315	1431800	1432174	hypothetical protein
CBO1317	1433826	1434953	hypothetical protein
CBO1319	1437105	1438550	Trk system potassium uptake protein TrkH
CBO1323	1444139	1444513	GntR family transcriptional regulator
CBO1325	1446393	1446980	rubrerythrin
CBO1326	1447275	1448729	cell wall hydrolase
CBO1328	1451384	1452031	UDP pyrophosphate synthetase-like protein
CBO1330	1453222	1453635	transcriptional regulator
CBO1331	1453844	1455730	oxidoreductase, acetyl-CoA synthase subunit
CBO1332	1455877	1456320	nitrate reductase, iron-sulfur subunit
CBO1333	1456519	1457745	nitrate reductase, NADH oxidase subunit
CBO1335	1459149	1459871	Crp/Fnr family transcriptional regulator
CBO1336	1459952	1461151	major facilitator superfamily protein
CBO1338	1462390	1463388	iron chelate uptake ABC transporter FeCT family,permease
CBO1341	1464783	1466060	(Fe-S)-binding protein
CBO1350	1473638	1475509	hypothetical protein
CBO1353	1477512	1477784	hypothetical protein
CBO1355	1480464	1482053	peptide/opine/nickel uptake ABC transporter PepTfamily, substrate-binding protein
CBO1356	1482050	1483006	peptide/opine/nickel uptake ABC transporter permease
CBO1360	1485958	1486470	lipoprotein
CBO1362	1487551	1488204	manganese/zinc/iron chelate uptake ABC transporter ATP-binding protein
CBO1363	1488425	1489363	manganese/zinc/iron chelate uptake ABC transporter Mzt family, substrate-binding protein
CBO1366	1492472	1493020	signaling protein
CBO1368	1493954	1494961	lipoprotein
CBO1369	1494933	1495694	ABC transporter permease
CBO1370	1495835	1496434	ABC transporter ATP-binding protein
CBO1373	1497156	1497629	hypothetical protein
CBO1374	1497756	1499027	zinc carboxypeptidase
CBO1378	1502167	1502544	hypothetical protein
CBO1379	1502557	1503714	sensor histidine kinase
CBO1380	1503897	1504550	LuxR family transcriptional regulator

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO1385	1509153	1509704	nitroreductase
CBO1386	1510262	1510753	hypothetical protein
CBO1388	1511391	1512854	aminoacyl-histidine dipeptidase
CBO1390	1514203	1515039	pyridoxamine kinase
CBO1392	1517266	1519155	peptide ABC transporter Pep4E family, permease
CBO1393	1519328	1519825	hypothetical protein
CBO1395	1520968	1521945	sensor histidine kinase
CBO1397	1522749	1524557	peptide ABC transporter Pep4E family, permease
CBO1400	1527112	1528290	aminotransferase, classes I and II
CBO1401	1528378	1529043	amino acid ABC transporter permease
CBO1402	1529062	1529799	polar amino acid ABC transporter ATP-binding protein
CBO1406	1532888	1533235	hypothetical protein
CBO1408	1535118	1535342	hypothetical protein
CBO1410	1537174	1538655	nicotinate phosphoribosyltransferase
CBO1412	1539920	1540579	amino acid ABC transporter permease
CBO1413	1540710	1541258	hypothetical protein
CBO1414	1541353	1543365	ATP-dependent DNA helicase
CBO1429	1551748	1552050	hypothetical protein
CBO1430	1552253	1554481	cell division protein FtsA
CBO1431	1554701	1555879	amidohydrolase
CBO1433	1557266	1558219	pseudouridine synthase
CBO1434	1558251	1559858	stage V sporulation protein B
CBO1435	1559895	1561400	polysaccharide biosynthesis family protein
CBO1436	1561733	1562296	hypothetical protein
CBO1437	1562317	1562724	hypothetical protein
CBO1447	1578828	1579664	nicotinate-nucleotide pyrophosphorylase
CBO1447A	1579645	1580946	L-aspartate oxidase
CBO1449	1582235	1583161	hypothetical protein
CBO1450	1583258	1583896	hemolysin III
CBO1451	1584017	1585114	GTP-dependent nucleic acid-binding protein EngD
CBO1454	1586357	1587286	S-adenosyl-methyltransferase MraW
CBO1459	1593198	1594151	phospho-N-acetylmuramoyl-pentapeptide- transferase
CBO1460	1594230	1595342	cell cycle protein, FtsW/RodA/SpoVE family
CBO1462	1596442	1597155	hypothetical protein
CBO1464	1597540	1598271	hypothetical protein
CBO1465	1598364	1599035	hypothetical protein
CBO1466	1599056	1599508	hypothetical protein
CBO1467	1599525	1599800	hypothetical protein
CBO1469	1600587	1601246	cell-division initiation protein DivIVA
CBO1470	1601654	1602742	3-dehydroquinate synthase
CBO1471	1602743	1603435	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase
CBO1472	1603556	1604356	3-deoxy-7-phosphoheptulonate synthase
CBO1473	1604490	1605818	3-phosphoshikimate 1-carboxyvinyltransferase
CBO1474	1606035	1606658	RNA polymerase-binding protein DksA
CBO1476	1607509	1608426	pseudouridine synthase
CBO1479	1609493	1610836	uracil permease
CBO1480	1611426	1612187	exodeoxyribonuclease III
CBO1481	1612410	1613552	methyltransferase
CBO1486	1617055	1617615	hypothetical protein
CBO1487	1617997	1618920	patatin-like phospholipase
CBO1488	1619220	1619375	hypothetical protein
CBO1490	1620788	1622185	Na+/H+ antiporter
CBO1492	1624147	1625076	CobW/P47K family protein
CBO1493	1625099	1625971	methylcobalamin:coenzyme M methyltransferase
CBO1494	1625997	1626659	dimethylamine corrinoid protein
CBO1495	1626744	1627766	methylcobalamin:coenzyme M methyltransferase
CBO1499	1629805	1630767	D-methionine ABC transporter ATP-binding protein
CBO1501	1631569	1632387	lipoprotein, NLPA family
CBO1503	1633625	1634644	FAD/FMN-binding oxidoreductase

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO1504	1634973	1635995	asparagine synthetase AsnA
CBO1506	1638024	1639883	DNA topoisomerase
CBO1509	1643334	1643825	hypothetical protein
CBO1514	1647969	1648520	hypothetical protein
CBO1516	1649623	1650516	carbohydrate uptake ABC transporter CUT1 family,permease
CBO1517	1650552	1651376	carbohydrate uptake ABC transporter CUT1 family,permease
CBO1519	1653007	1653963	L-lactate dehydrogenase
CBO1520	1654493	1655203	GntR family transcriptional regulator
CBO1522	1656372	1657319	C4-dicarboxylate transporter/malic acid transport protein
CBO1525	1660488	1660835	hypothetical protein
CBO1528	1663975	1664517	hypothetical protein
CBO1532	1665808	1666278	methylated-DNA--protein-cysteine methyltransferase
CBO1534	1667687	1668895	hypothetical protein
CBO1535	1668920	1669750	hypothetical protein
CBO1541	1675185	1675544	iron-dependent repressor
CBO1551	1684283	1685227	DeoR family transcriptional regulator
CBO1552	1685851	1686846	2-hydroxyglutaryl-CoA dehydratase subunit
CBO1554	1687742	1688173	6-pyruvoyl tetrahydrobiopterin synthase
CBO1555	1688175	1688840	radical SAM domain protein
CBO1556	1688844	1689434	GTP cyclohydrolase I
CBO1557	1689618	1690277	hypothetical protein
CBO1563	1694268	1694906	deoxyribose-phosphate aldolase
CBO1577	1706132	1706659	hypothetical protein
CBO1578	1706690	1707580	LysR family transcriptional regulator
CBO1579	1707844	1709094	sodium:alanine symporter family protein
CBO1581	1711028	1712284	hypothetical protein
CBO1582	1712582	1713298	antibiotic ABC transporter ATP-binding protein
CBO1587	1717488	1718714	arginine deiminase
CBO1588	1718839	1720257	arginine/ornithine antiporter
CBO1590	1721498	1722094	hypothetical protein
CBO1596	1727629	1728444	RelA/SpoT domain-containing protein
CBO1601	1734575	1735576	LacI family transcriptional regulator
CBO1602	1735880	1736827	glucokinase
CBO1604	1737074	1738738	glycosyl hydrolase family protein
CBO1605	1738780	1740438	PTS system, glucose/glucoside family, IIBC component
CBO1607	1741458	1742081	hypothetical protein
CBO1614	1748904	1750130	major facilitator transporter
CBO1619	1753552	1753743	hypothetical protein
CBO1621	1758029	1758916	5,10-methylenetetrahydrofolate reductase
CBO1622	1759080	1759769	hypothetical protein
CBO1623	1759759	1762137	5-methyltetrahydrofolate--homocysteine methyltransferase
CBO1630	1768164	1769147	TPP-dependent acetoin dehydrogenase complex, E1 component subunit alpha
CBO1631	1769150	1770121	TPP-dependent acetoin dehydrogenase complex, E1 component subunit beta
CBO1633	1771205	1772515	TPP-dependent acetoin dehydrogenase complex, E2 component, dihydrolipoamide acetyltra
CBO1634	1772528	1773919	TPP-dependent acetoin dehydrogenase complex, E3 component, dihydrolipoamide dehydro
CBO1637	1776428	1778014	glycine betaine/L-proline ABC transporter permease/substrate-binding protein
CBO1638	1778829	1780310	threonine synthase
CBO1641	1782882	1784159	homoserine dehydrogenase
CBO1644	1788660	1789547	ferredoxin-NADP reductase
CBO1645	1789547	1790929	oxidoreductase
CBO1647	1793129	1793830	pentapeptide repeat-containing protein
CBO1659	1800539	1801093	hypothetical protein
CBO1660	1801531	1802700	alanine racemase
CBO1663	1806261	1807094	hypothetical protein
CBO1664	1807311	1808195	oxidoreductase, short chain dehydrogenase/reductase
CBO1667	1811590	1812603	hypothetical protein
CBO1757	1866035	1866622	DNA polymerase III subunit epsilon
CBO1758	1866917	1867867	hypothetical protein
CBO1759	1868139	1868774	UbiE/COQ5 family methyltransferase

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO1762	1871048	1871611	TetR family transcriptional regulator
CBO1763	1871811	1872716	fructose-1-phosphate kinase
CBO1773	1880679	1882682	methyl-accepting chemotaxis protein
CBO1774	1882963	1884543	drug resistance ABC transporter ATP-binding protein
CBO1776	1885911	1886765	sulfite/nitrite reductase
CBO1777	1886905	1887798	chemotaxis protein CheV
CBO1778	1888167	1889432	ATP-dependent RNA helicase, DEAD/DEAH box family
CBO1779	1889755	1890591	RNA-binding protein
CBO1780	1890770	1892215	ATP-dependent RNA helicase DeaD
CBO1782	1897225	1901304	hypothetical protein
CBO1786	1903685	1904539	chemotaxis sensory transduction protein
CBO1788	1905183	1906541	hypothetical protein
CBO1795	1911268	1911873	LexA repressor
CBO1796	1912070	1913353	aluminum resistance protein
CBO1797	1913454	1913711	RNA chaperone Hfq
CBO1798	1913797	1914732	tRNA delta(2)-isopentenylpyrophosphate transferase
CBO1799	1914748	1916517	DNA mismatch repair protein
CBO1801	1919541	1920893	(dimethylallyl)adenosine tRNA methylthiotransferase
CBO1802	1921365	1922555	aminotransferase, classes I and II
CBO1803	1922640	1922801	hypothetical protein
CBO1804	1922856	1924046	aspartate aminotransferase
CBO1806	1925025	1926083	spore coat protein
CBO1807	1926331	1928247	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
CBO1810	1929411	1930634	pyridine nucleotide-disulfide oxidoreductase
CBO1811	1930929	1932194	glutamate dehydrogenase
CBO1812	1932429	1933313	transcriptional regulator
CBO1813	1933681	1934391	ribosomal large subunit pseudouridine synthase B
CBO1814	1934854	1936143	hypothetical protein
CBO1816	1937816	1939219	glycosyltransferase
CBO1817	1939580	1939780	hypothetical protein
CBO1819	1942218	1942904	cytochrome c-type biogenesis protein
CBO1820	1942906	1943337	thioredoxin family protein
CBO1822	1944419	1944973	hypothetical protein
CBO1825	1946943	1947287	4-carboxymuconolactone decarboxylase
CBO1826	1947425	1947940	GNAT family acetyltransferase
CBO1827	1948125	1948619	lipoprotein
CBO1828	1948851	1949162	hypothetical protein
CBO1828A	1949143	1949376	hypothetical protein
CBO1830	1949755	1951479	hydrogenase
CBO1844	1965696	1965869	hypothetical protein
CBO1847	1969454	1971187	[Fe] hydrogenase
CBO1848	1971207	1973102	[Fe] hydrogenase, electron-transfer subunit
CBO1851	1975515	1976393	ADP-ribosylglycohydrolase
CBO1852	1976634	1977728	hypothetical protein
CBO1853	1977838	1978974	D-alanyl-D-alanine carboxypeptidase
CBO1855	1979738	1980178	sporulation protein YtfJ
CBO1856	1980239	1980820	segregation and condensation protein B
CBO1857	1980813	1981559	segregation and condensation protein A
CBO1862	1985764	1986012	hypothetical protein
CBO1863	1986119	1987423	pyrimidine-nucleoside phosphorylase
CBO1864	1987709	1988524	purine nucleoside phosphorylase
CBO1865A	1989489	1989716	hypothetical protein
CBO1866	1989707	1990360	sporulation protein
CBO1867	1990565	1991101	hydrolase, NUDIX family
CBO1872	1998002	1998823	sporulation transcription factor Spo0A
CBO1875	2001824	2003050	homoserine dehydrogenase
CBO1876	2003090	2004370	O-acetylhomoserine aminocarboxypropyltransferase/cysteine synthase
CBO1878	2006385	2006837	arginine repressor ArgR
CBO1881	2008527	2010395	1-deoxy-D-xylulose-5-phosphate synthase

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO1885	2013101	2014306	exodeoxyribonuclease VII large subunit
CBO1886	2014377	2014808	transcription antitermination protein NusB
CBO1887	2014947	2015342	alkaline shock protein
CBO1888	2015424	2015948	stage III sporulation protein AH
CBO1890	2016744	2017334	stage III sporulation protein AF
CBO1891	2017370	2018533	stage III sporulation protein AE
CBO1892	2018542	2018925	stage III sporulation protein AD
CBO1893	2018952	2019149	stage III sporulation protein AC
CBO1894	2019192	2019710	stage III sporulation protein SpoAB
CBO1895	2019747	2020670	stage III sporulation protein AA
CBO1896	2020865	2021284	hypothetical protein
CBO1897	2021325	2021882	elongation factor P
CBO1898	2021971	2022414	3-dehydroquinate dehydratase
CBO1901	2023715	2024161	hypothetical protein
CBO1902	2024177	2024893	hypothetical protein
CBO1904	2025319	2025759	hypothetical protein
CBO1908	2027627	2028817	type II secretion system protein F
CBO1911	2030864	2031937	lipoprotein
CBO1913	2033013	2034743	fibronectin-binding protein
CBO1914	2034812	2035552	tRNA pseudouridine synthase A
CBO1916	2036092	2036469	HIT family protein
CBO1917	2036980	2038383	glutamate decarboxylase
CBO1918	2038816	2039157	hypothetical protein
CBO1920	2041841	2043424	clostripain
CBO1921	2043895	2044551	exported protein
CBO1921b	2044553	2045254	DNA-binding response regulator
CBO1927	2051571	2052347	hypothetical protein
CBO1929	2053354	2054097	ABC transporter ATP-binding protein
CBO1930	2054687	2056549	ABC transporter ATP-binding protein/permease
CBO1931	2056546	2058873	ABC transporter permease/ATP-binding protein
CBO1932	2059277	2059720	MarR family transcriptional regulator
CBO1934	2060717	2060944	molybdopterin converting factor subunit 1
CBO1935	2061097	2062914	aldehyde ferredoxin oxidoreductase
CBO1938	2065268	2065708	MOSC domain-containing protein
CBO1941	2067211	2069115	LysR family transcriptional regulator
CBO1942	2069127	2070353	molybdopterin biosynthesis protein MoeA
CBO1943	2070377	2070865	molybdopterin biosynthesis protein Mog
CBO1944	2070960	2071652	tungstate ABC transporter ATP-binding protein
CBO1945	2071649	2072320	tungstate ABC transporter permease
CBO1946	2072329	2073180	tungstate ABC transporter tungstate-binding protein
CBO1947	2073399	2074721	drug/sodium antiporter
CBO1951	2076811	2078274	ABC transporter permease
CBO1952	2079073	2080830	ABC transporter permease/ATP-binding protein
CBO1953	2080827	2082557	ABC transporter ATP-binding/permease
CBO1954	2082672	2083493	MerR family transcriptional regulator
CBO1955	2083912	2084538	NADPH-dependent FMN reductase
CBO1966	2097640	2099859	secreted protein
CBO1968	2101489	2102154	DNA-binding response regulator
CBO1969	2102654	2103469	hypothetical protein
CBO1970	2103495	2104220	ABC transporter permease
CBO1972	2105701	2106267	ABC transporter permease
CBO1976	2110541	2111632	spore germination protein
CBO1977	2111616	2112734	spore germination protein
CBO1979	2113913	2114341	hypothetical protein
CBO1982	2114981	2115961	radical SAM domain protein
CBO1987	2123615	2125042	PTS system, trehalose-specific IIBC component
CBO1988	2125124	2126791	alpha,alpha-phosphotrehalase
CBO1989	2126824	2127534	trehalose operon repressor
CBO1991	2128609	2130006	PTS system fructose family transporter subunit IIBC

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO1992	2130007	2130930	kinase, pfkB family
CBO1993	2130950	2131810	tagatose-bisphosphate aldolase
CBO1994	2132010	2132747	transcriptional regulator
CBO1997	2134343	2135212	radical SAM domain protein
CBO1999	2135788	2136879	glycerol dehydrogenase
CBO2000	2137371	2139020	methyl-accepting chemotaxis protein
CBO2003	2140610	2140888	hypothetical protein
CBO2005	2142543	2143088	adenine phosphoribosyltransferase
CBO2006	2143104	2144168	xanthine/uracil permease
CBO2009	2146957	2148732	sigma-54 dependent transcriptional regulator
CBO2010	2149198	2150328	ATP-dependent RNA helicase, DEAD/DEAH box family
CBO2012	2151548	2152894	GntR family transcriptional regulator
CBO2015	2154249	2155400	aspartate aminotransferase
CBO2020	2159434	2160573	class V aminotransferase
CBO2024	2162319	2162540	hypothetical protein
CBO2029	2166279	2167625	proton-dependent oligopeptide transporter
CBO2030	2168134	2169222	Xaa-Pro dipeptidase
CBO2041	2181659	2182894	hypothetical protein
CBO2075	2211857	2212981	radical SAM domain protein
CBO2106	2246993	2247283	microcompartments family protein
CBO2107	2247357	2248007	propanediol utilization protein PduL-like protein
CBO2108	2248007	2249605	acetaldehyde dehydrogenase
CBO2109	2249640	2250437	propanediol/ethanolamine utilization protein
CBO2110	2250458	2251120	ethanolamine utilization protein EutQ-like protein
CBO2111	2251319	2251582	polyhedral organelle formation protein
CBO2112	2251579	2251881	hypothetical protein
CBO2113	2251959	2252804	ethanolamine utilization protein EutJ family protein
CBO2114	2252786	2253514	hypothetical protein
CBO2119	2258064	2259557	aldehyde dehydrogenase
CBO2121	2259927	2260223	microcompartments family protein
CBO2123	2260527	2260793	hypothetical protein
CBO2124	2260780	2261892	hypothetical protein
CBO2125	2262378	2263208	MerR family transcriptional regulator
CBO2126	2263325	2264467	iron-containing alcohol dehydrogenase
CBO2127	2264980	2266281	transmembrane MATE family transporter protein
CBO2135	2274479	2275498	sensor histidine kinase
CBO2139	2279295	2280170	ABC transporter ATP-binding protein
CBO2142	2283872	2284363	hypothetical protein
CBO2146	2288568	2290571	ABC transporter permease
CBO2148	2292105	2292458	hypothetical protein
CBO2150	2295089	2296453	PTS system ascorbate-specific transporter subunit IIC
CBO2151	2296466	2296750	PTS system L-ascorbate family transporter subunit IIB
CBO2152	2296779	2297216	PTS system L-ascorbate family transporter subunit IIA
CBO2153	2297240	2297956	translaldolase
CBO2155	2298933	2299667	hypothetical protein
CBO2162	2304858	2306264	proton-dependent oligopeptide family transport protein
CBO2165	2308305	2308826	hypothetical protein
CBO2166	2310725	2312722	2-enoate reductase
CBO2167	2312819	2313616	MerR family transcriptional regulator
CBO2191	2339703	2341022	amino acid permease
CBO2194	2344083	2345132	electron transfer flavoprotein subunit alpha
CBO2195	2345151	2345936	electron transfer flavoprotein subunit beta
CBO2196	2345975	2347108	acyl-CoA dehydrogenase
CBO2197	2347373	2348500	subunit of oxygen-sensitive 2-hydroxyisocaproyl-CoA dehydratase
CBO2198	2348500	2349732	subunit of oxygen-sensitive 2-hydroxyisocaproyl-CoA dehydratase
CBO2199	2349795	2351009	isocaprenoyl-CoA:2-hydroxyisocaproate CoA-transferase
CBO2221	2379184	2379540	nitrogenase iron-molybdenum cofactor protein
CBO2224	2381327	2381686	dinitrogenase iron-molybdenum cofactor family protein
CBO2233	2392460	2393488	low specificity L-threonine aldolase

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO2236	2396107	2396904	oxidoreductase, short chain dehydrogenase/reductase
CBO2238	2397426	2398604	sodium extrusion protein NatB
CBO2239	2398618	2399337	sodium extrusion ATP-binding protein NatA
CBO2243	2401913	2402983	uroporphyrinogen decarboxylase
CBO2244	2403331	2404779	oxidoreductase, FAD/[2Fe-2S]-binding
CBO2245	2405127	2405945	MerR family transcriptional regulator
CBO2246	2406153	2407556	amino acid permease
CBO2256	2413523	2414251	metallo-beta-lactamase
CBO2259	2415679	2416635	biotin synthase
CBO2262	2418214	2418810	hypothetical protein
CBO2266	2423261	2423530	hypothetical protein
CBO2277	2432526	2433122	Mn/Fe superoxide dismutase
CBO2281	2438518	2439252	hypothetical protein
CBO2283	2440027	2440947	antibiotic ABC transporter ATP-binding protein
CBO2284	2441149	2442060	sensor histidine kinase
CBO2286	2443391	2444152	exonuclease
CBO2290	2446163	2446402	DNA-binding protein
CBO2396	2524131	2525561	adenylosuccinate lyase
CBO2397	2525804	2526034	hypothetical protein
CBO2398	2526062	2526319	phosphocarrier protein HPr
CBO2400	2527768	2528124	hypothetical protein
CBO2401	2528121	2529404	hypothetical protein
CBO2402	2529423	2530289	phosphoribose diphosphate:decaprenyl-phosphate phosphoribosyltransferase
CBO2403	2530452	2530712	stage V sporulation protein S
CBO2404	2530851	2532392	phosphodiesterase
CBO2406	2533833	2534426	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase
CBO2407	2534401	2535738	MiaB family RNA modification protein
CBO2408	2535833	2538109	DNA translocase FtsK/SpoIIIE
CBO2410	2539530	2540735	aspartate kinase
CBO2411	2540751	2541023	PRC-barrel domain-containing protein
CBO2412	2541100	2542401	peptidase, M16 family
CBO2416	2546277	2547146	tRNA pseudouridine synthase B
CBO2417	2547147	2548118	DHH family protein
CBO2417A	2548099	2548461	ribosome-binding factor A
CBO2418	2548492	2550558	translation initiation factor IF-2
CBO2420	2551212	2552585	transcription elongation factor NusA
CBO2421	2552602	2553063	hypothetical protein
CBO2423	2553368	2557666	DNA polymerase III PolC
CBO2425	2559135	2560145	membrane-associated zinc metalloprotease
CBO2426	2560190	2561347	1-deoxy-D-xylulose 5-phosphate reductoisomerase
CBO2431	2563841	2564599	UDP pyrophosphate synthase
CBO2433	2565235	2565951	uridylate kinase
CBO2434	2566326	2567249	elongation factor Ts
CBO2435	2567337	2568038	30S ribosomal protein S2
CBO2436	2568275	2569051	transcriptional repressor CodY
CBO2441	2575010	2575819	ribonuclease HII
CBO2442	2576116	2576994	ribosomal biogenesis GTPase
CBO2443	2577299	2577823	signal peptidase I
CBO2444	2577882	2578226	50S ribosomal protein L19
CBO2445	2578377	2579099	tRNA (guanine-N(1)-)-methyltransferase
CBO2446	2579090	2579584	16S rRNA-processing protein RimM
CBO2447	2579697	2579924	hypothetical protein
CBO2448	2579945	2580193	30S ribosomal protein S16
CBO2449	2580224	2581573	signal recognition particle protein
CBO2451	2582142	2583053	signal recognition particle-docking protein FtsY
CBO2453	2587061	2588152	radical SAM domain protein
CBO2454	2588145	2588858	ribonuclease III
CBO2455	2588961	2589203	acyl carrier protein
CBO2456	2589272	2590279	glycerol-3-phosphate acyltransferase PlsX

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO2457	2590369	2590560	50S ribosomal protein L32
CBO2458	2591182	2592375	acetate kinase
CBO2460	2593847	2594725	hypothetical protein
CBO2464	2596125	2598239	proline reductase
CBO2472	2607869	2609314	amino acid permease
CBO2474	2610771	2611778	proline racemase
CBO2479	2614194	2614469	hypothetical protein
CBO2484	2619269	2620573	electron transferring subunit of proline reductase
CBO2485	2621166	2621693	hypothetical protein
CBO2486	2621698	2622177	hypothetical protein
CBO2490	2624598	2626352	sensory box sigma-54 dependent transcriptional regulator
CBO2491	2626688	2627917	hypothetical protein
CBO2493	2629144	2629668	hypothetical protein
CBO2494	2629681	2630175	phosphopantetheine adenylyltransferase
CBO2495	2630177	2630734	methyltransferase
CBO2496	2630996	2633035	ATP-dependent DNA helicase RecG
CBO2497	2633385	2635010	DAK2 domain-containing protein
CBO2498	2635025	2635375	hypothetical protein
CBO2500A	2636012	2636167	hypothetical protein
CBO2502	2637191	2637841	ribulose-phosphate 3-epimerase
CBO2503	2637835	2638713	ribosome-associated GTPase
CBO2505	2640841	2641599	protein phosphatase
CBO2510	2645990	2646433	peptide deformylase
CBO2513	2650395	2651024	guanylate kinase
CBO2514	2651024	2651299	hypothetical protein
CBO2516	2652387	2653388	type II L-asparaginase
CBO2517	2654828	2656309	stage IV sporulation protein A
CBO2519	2658076	2659395	GTP-binding protein EngA
CBO2520	2659398	2660732	hypothetical protein
CBO2521	2661184	2661837	phosphate transporter PhoU
CBO2522	2661867	2662616	phosphate transporter ATP-binding protein
CBO2523	2662629	2663513	phosphate ABC transporter permease
CBO2524	2663513	2664436	phosphate ABC transporter permease
CBO2525	2664644	2665516	phosphate ABC transporter phosphate-binding protein
CBO2527	2667484	2668182	phosphate regulon DNA-binding response regulator
CBO2528	2668193	2668912	hypothetical protein
CBO2530	2669361	2669816	transcriptional regulator NrdR
CBO2532	2670357	2671130	sporulation sigma factor SigG
CBO2533	2671205	2671912	sporulation sigma factor SigE
CBO2535	2673034	2674143	cell division protein FtsZ
CBO2536	2674166	2675419	cell division protein FtsA
CBO2539	2677509	2678561	twitching motility protein
CBO2540	2678563	2679339	shikimate 5-dehydrogenase
CBO2554	2690007	2690627	uridine kinase
CBO2556	2691994	2692650	O-methyltransferase
CBO2559	2696312	2698015	RNA-metabolizing metallo-beta-lactamase
CBO2560	2698069	2698521	FUR family transcriptional regulator
CBO2561	2698643	2698897	hypothetical protein
CBO2562	2698911	2699324	Holliday junction resolvase-like protein
CBO2563	2699531	2699782	hypothetical protein
CBO2567	2704746	2705174	iron-sulfur cluster assembly protein
CBO2568	2705176	2706369	cysteine desulfurase
CBO2569	2706362	2706808	transcriptional regulator
CBO2570	2707028	2708278	recombination factor protein RarA
CBO2571	2708676	2709443	histidinol-phosphatase
CBO2573	2710826	2711692	DegV family protein
CBO2574	2712020	2712604	hypothetical protein
CBO2575	2712856	2713845	ornithine cyclodeaminase
CBO2578	2717790	2718443	hypothetical protein

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO2594	2738132	2739076	carbamate kinase
CBO2598	2742377	2742646	hypothetical protein
CBO2600	2743784	2744260	lipoprotein
CBO2602	2749613	2750836	hypothetical protein
CBO2603	2750990	2752045	hypothetical protein
CBO2605	2753506	2754909	aminopeptidase
CBO2606	2755032	2755967	magnesium/cobalt transporter CorA
CBO2607	2756393	2756947	5-formyltetrahydrofolate cyclo-ligase
CBO2614	2762607	2763281	glycosyl transferase family protein
CBO2618	2766342	2766740	hypothetical protein
CBO2620	2768911	2769417	hypothetical protein
CBO2623	2772103	2772762	V-type ATP synthase subunit D
CBO2624	2772782	2774167	V-type ATP synthase subunit B
CBO2625A	2775952	2776260	V-type ATP synthase subunit F
CBO2630	2780976	2782655	B12-binding Fe-S oxidoreductase
CBO2632	2783377	2784096	phospholipase C
CBO2635	2788220	2789863	manganese-dependent inorganic pyrophosphatase
CBO2637	2790985	2791776	flagellar basal body rod protein FlgG
CBO2638	2791831	2792598	flagellar basal body rod protein FlgG
CBO2641	2793525	2794253	RNA polymerase sigma factor for flagellar operonFliA
CBO2642	2794267	2794926	hypothetical protein
CBO2645	2797018	2799084	flagellar biosynthesis protein FlhA
CBO2647	2800975	2801244	flagellar biosynthesis protein FliQ
CBO2648	2801256	2802032	flagellar biosynthesis protein FliP
CBO2649	2802016	2802423	flagellar biosynthesis domain-containing protein
CBO2651	2803014	2803757	chemotaxis protein MotB
CBO2652	2803750	2804562	chemotaxis MotA protein
CBO2654	2805002	2806039	flagellar hook protein flgE
CBO2655	2806121	2806540	flagellar operon protein
CBO2657	2807382	2808770	flagellar hook-length control protein
CBO2658	2808776	2809213	flagellar protein FliJ
CBO2665	2814907	2815308	flagellar basal-body rod protein FlgB
CBO2668	2816878	2817246	hypothetical protein
CBO2669	2817786	2819108	argininosuccinate lyase
CBO2670	2819130	2820323	argininosuccinate synthase
CBO2672	2820618	2820956	hypothetical protein
CBO2674	2821503	2822936	cardiolipin synthetase
CBO2675	2823260	2824150	hypothetical protein
CBO2676	2824325	2824510	hypothetical protein
CBO2689	2837248	2837925	capsular polysaccharide biosynthsis protein
CBO2692	2839665	2839844	hypothetical protein
CBO2693	2839959	2840126	hypothetical protein
CBO2730	2882498	2883325	flagellin
CBO2734	2887373	2887771	flagellar protein FliS
CBO2735	2887799	2888098	hypothetical protein
CBO2736	2888153	2888509	flagellar protein FlaG
CBO2737	2888525	2888743	carbon storage regulator
CBO2738	2888743	2889174	flagellar assembly protein FliW
CBO2740	2890243	2892111	flagellar hook-associated protein FlgK
CBO2741	2892400	2892804	flgN family protein
CBO2742	2892804	2893085	regulator of flagellin synthesis FlgM
CBO2743	2893582	2894742	flagellar motor switch protein
CBO2744	2894735	2895730	flagellar motor switch protein FliM
CBO2745	2895758	2896153	chemotaxis protein CheW
CBO2746	2896253	2896612	chemotaxis protein CheY
CBO2747	2896640	2897239	chemotaxis protein CheC
CBO2748	2897265	2899340	chemotaxis protein CheA
CBO2749	2899353	2900123	chemotaxis protein CheR
CBO2750	2900138	2901202	protein-glutamate methylesterase CheB

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO2751	2901225	2901713	chemoreceptor glutamine deamidase CheD
CBO2752	2901730	2902176	chemotaxis protein CheW
CBO2754	2904478	2906433	PTS system glucose family transporter subunit IIABC
CBO2755	2906516	2907343	BglG family transcriptional antiterminator
CBO2756	2907592	2908458	branched-chain amino acid aminotransferase
CBO2758	2909450	2909851	hypothetical protein
CBO2762	2916368	2918476	sensory box histidine kinase
CBO2764	2920571	2921389	AP endonuclease, family 2
CBO2768	2927106	2927669	glycerol uptake operon antiterminator regulatoryprotein
CBO2769	2927718	2930588	DNA topoisomerase IV subunit A
CBO2772	2933548	2934720	transporter monovalent cation:proton antiporter-2 (CPA2) family
CBO2773	2934733	2935083	hypothetical protein
CBO2774	2935544	2935780	hypothetical protein
CBO2775	2935800	2936693	polysaccharide deacetylase
CBO2776	2936772	2938061	hypothetical protein
CBO2778	2939426	2939842	flavodoxin
CBO2780	2941209	2942630	MATE efflux family protein
CBO2784	2947271	2948767	glycerol kinase
CBO2788	2951530	2952606	ABC transporter permease
CBO2790	2953492	2953872	GntR family transcriptional regulator
CBO2793	2957454	2958044	membrane-spanning protein
CBO2794	2958348	2959580	aminopeptidase
CBO2798	2963737	2963955	hypothetical protein
CBO2803	2968107	2968274	hypothetical protein
CBO2806	2971113	2971523	universal stress protein family protein
CBO2808	2973057	2973974	glutaminase
CBO2812	2978651	2979451	hypothetical protein
CBO2818	2987038	2988087	DNA polymerase IV
CBO2819	2988552	2989787	dihydropyrimidine dehydrogenase
CBO2821	2991550	2992098	pyrimidine operon regulatory protein/uracil phosphoribosyltransferase
CBO2824	2994349	2996607	excinuclease ABC subunit A
CBO2827	2998282	2998680	hypothetical protein
CBO2829	3001344	3001781	MutT/nudix family protein
CBO2830	3002321	3002788	FUR family transcriptional regulator
CBO2831	3002921	3004918	hypothetical protein
CBO2833	3007452	3008186	glucosamine-6-phosphate deaminase
CBO2834	3008254	3009390	N-acetylglucosamine-6-phosphate deacetylase
CBO2837	3011655	3012134	PTS system glucose subfamily transporter subunitIIA
CBO2839	3012881	3014323	PTS system, N-acetylglucosamine-specific, IIBC component
CBO2840	3014803	3016173	MATE efflux family protein
CBO2841	3016338	3017732	MATE efflux family protein
CBO2842	3017966	3019165	metallo-beta-lactamase/flavodoxin
CBO2843	3019224	3021131	acyl-CoA dehydrogenase
CBO2857	3035493	3036527	chloramphenicol/florfenicol resistance protein
CBO2859	3038632	3040002	23S rRNA (uracil-5-)-methyltransferase RumA
CBO2860	3040274	3040720	hypothetical protein
CBO2861	3041051	3042910	sulfatase
CBO2862	3042936	3043688	hypothetical protein
CBO2863	3044098	3044562	6,7-dimethyl-8-ribityllumazine synthase
CBO2864	3044594	3045799	bifunctional riboflavin biosynthesis protein RibAB
CBO2866	3046774	3047871	riboflavin biosynthesis protein RibD
CBO2868	3050092	3050985	metallo-beta-lactamase
CBO2871	3054611	3055015	hypothetical protein
CBO2872	3055343	3056584	phosphoribosylamine--glycine ligase
CBO2874	3058446	3059063	phosphoribosylglycinamide formyltransferase
CBO2875	3059191	3060195	phosphoribosylaminoimidazole synthetase
CBO2877	3061787	3062491	phosphoribosylaminoimidazole-succinocarboxamide synthase
CBO2878	3062491	3062970	phosphoribosylaminoimidazole carboxylase catalytic subunit
CBO2881	3063646	3066216	xanthine dehydrogenase, molybdopterin-binding subunit

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO2882	3066576	3067793	peptidase
CBO2886	3073769	3075163	phenylhydantoinase
CBO2887	3075213	3076619	xanthine/uracil permease
CBO2890	3080263	3081000	hypothetical protein
CBO2891	3081015	3081629	hypothetical protein
CBO2892	3081610	3082458	hypothetical protein
CBO2893	3082440	3083249	xanthine dehydrogenase accessory factor
CBO2894	3083275	3083820	hypothetical protein
CBO2895	3083817	3084785	iron ABC transporter permease
CBO2896	3084782	3085576	iron ABC transporter ATP-binding protein
CBO2899	3087284	3087856	lipase/acylhydrolase, GDSL family
CBO2911	3102194	3103507	thiamine biosynthesis protein ThiC
CBO2914	3105357	3105638	hypothetical protein
CBO2918	3108106	3108477	GntR family transcriptional regulator
CBO2922	3111145	3111888	(Fe-S)-binding protein
CBO2924	3112759	3113202	cell wall hydrolase
CBO2925	3113225	3113782	hypothetical protein
CBO2927	3114506	3116668	phage infection protein
CBO2928	3116976	3117728	glycerophosphoryl diester phosphodiesterase
CBO2929	3118107	3119069	TPR domain protein
CBO2934	3122852	3123586	hypothetical protein
CBO2935	3123725	3124528	hypothetical protein
CBO2936	3124519	3125208	hypothetical protein
CBO2937	3125284	3126171	hypothetical protein
CBO2938	3126246	3127331	RNA polymerase sigma factor RpoD
CBO2939	3127356	3129128	DNA primase
CBO2940	3129411	3130439	deoxyguanosinetriphosphate triphosphohydrolase-like protein
CBO2941	3130581	3131651	spore coat protein-like protein
CBO2942	3131704	3134337	pyruvate phosphate dikinase
CBO2943	3134417	3135058	hypothetical protein
CBO2946	3136967	3137857	GTP-binding protein Era
CBO2947	3137873	3138271	cytidine deaminase
CBO2948	3138355	3139053	diacylglycerol kinase/PAP2 family protein
CBO2949	3139092	3139592	metalloprotease
CBO2950	3139589	3141676	hydrolase
CBO2951	3141699	3142859	stage IV sporulation protein
CBO2952	3143259	3143705	GatB/Yqey domain protein
CBO2953	3143748	3143924	30S ribosomal protein S21
CBO2955	3144462	3145760	MiaB family RNA modification protein
CBO2956	3145760	3146518	16S ribosomal RNA methyltransferase RsmE
CBO2957	3146616	3147554	50S ribosomal protein L11 methyltransferase
CBO2959	3149020	3150891	molecular chaperone DnaK
CBO2960	3150946	3151590	heat shock protein GrpE
CBO2961	3151617	3152648	heat-inducible transcription repressor
CBO2963	3154018	3155826	GTP-binding protein LepA
CBO2964	3155967	3156362	hypothetical protein
CBO2965	3156433	3157524	stage II sporulation protein
CBO2966	3157709	3158683	germination protease
CBO2967	3158893	3159159	30S ribosomal protein S20
CBO2969	3160254	3162041	ComEC/Rec2 family protein
CBO2970	3162053	3162283	hypothetical protein
CBO2975	3168726	3170633	selenocysteine-specific translation elongation factor
CBO2976	3170652	3172040	selenocysteine synthase
CBO2978	3173292	3173900	comE operon protein 1
CBO2981	3176672	3177052	endoribonuclease L-PSP
CBO2983	3178353	3178922	HDIG domain protein
CBO2985	3179853	3180146	RNA-binding protein
CBO2986	3180173	3181447	GTPase ObgE
CBO2987	3181826	3182128	50S ribosomal protein L27

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO2989	3182460	3182774	50S ribosomal protein L21
CBO2990	3182898	3184337	ribonuclease, Rne/Rng family
CBO2992	3185186	3187039	radical SAM domain protein
CBO2995	3189160	3190278	rod shape-determining protein RodA
CBO2996	3190377	3190640	cell division topological specificity factor MinE
CBO2997	3190654	3191451	septum site-determining protein MinD
CBO2998	3191469	3192131	septum formation inhibitor
CBO2999	3192273	3195047	penicillin-binding protein
CBO3001	3195569	3196414	rod shape-determining protein MreC
CBO3002	3196421	3197431	rod shape-determining protein MreB
CBO3003	3197460	3198146	DNA repair protein RadC
CBO3005	3198761	3199018	hypothetical protein
CBO3007	3200055	3200594	hypothetical protein
CBO3009	3201020	3201574	hypothetical protein
CBO3010	3201633	3202070	hypothetical protein
CBO3012	3202805	3203350	transglycosylase SLT domain protein
CBO3014	3204018	3206645	DNA polymerase I
CBO3015	3206663	3207787	chorismate synthase
CBO3050	3239071	3239190	hypothetical protein
CBO3051	3239247	3239537	hypothetical protein
CBO3052	3239678	3239806	hypothetical protein
CBO3053	3239978	3240406	sigma-E factor regulator, RseC/MucC family
CBO3054	3240441	3242222	aspartyl-tRNA synthetase
CBO3055	3242263	3243510	histidyl-tRNA synthetase
CBO3057	3244997	3245596	metallo-beta-lactamase
CBO3058	3245657	3246106	D-tyrosyl-tRNA(Tyr) deacylase
CBO3059	3246149	3248320	GTP pyrophosphokinase
CBO3060	3248444	3248962	adenine phosphoribosyltransferase
CBO3061	3249071	3249949	DHH domain-containing protein
CBO3062	3250178	3251041	preprotein translocase subunit SecF
CBO3063	3251043	3252311	preprotein translocase subunit SecD
CBO3065	3253929	3254066	hypothetical protein
CBO3067	3254579	3254914	preprotein translocase subunit YajC
CBO3068	3255094	3256224	queuine tRNA-ribosyltransferase
CBO3069	3256280	3257305	S-adenosylmethionine--tRNA ribosyltransferase-isomerase
CBO3070	3257339	3258367	Holliday junction DNA helicase RuvB
CBO3071	3258387	3258980	Holliday junction DNA helicase RuvA
CBO3072	3259162	3259773	hypothetical protein
CBO3073	3260150	3260890	hypothetical protein
CBO3074	3261041	3261691	hypothetical protein
CBO3075	3261707	3262369	PBS lyase HEAT-like repeat protein
CBO3076	3262457	3263875	GntR family transcriptional regulator
CBO3079	3265598	3265756	rubredoxin
CBO3080	3265885	3267669	GTP-binding protein
CBO3083	3269197	3271689	penicillin-binding protein, 1A family
CBO3087	3273778	3274533	sporulation sigma factor SigF
CBO3088	3274552	3274974	anti-sigma F factor
CBO3089	3274991	3275326	anti-sigma F factor antagonist
CBO3090	3275627	3277372	modification methylase
CBO3091	3277427	3278401	polysaccharide deacetylase
CBO3092	3278490	3279206	WecB/TagA/CpsF family glycosyl transferase
CBO3095	3282030	3283586	integral membrane protein MviN
CBO3096	3283605	3284708	glycoside hydrolase family protein
CBO3107	3296631	3297740	oxidoreductase, NAD-binding Rossmann fold family
CBO3108	3297803	3299119	UDP-glucose/GDP-mannose dehydrogenase
CBO3109	3299249	3301114	exopolysaccharide biosynthesis protein
CBO3112	3307383	3308279	UTP-glucose-1-phosphate uridylyltransferase
CBO3115	3312263	3312799	50S ribosomal protein L25
CBO3116	3313006	3314466	Orn/Lys/Arg decarboxylase

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO3117	3314651	3315484	hypothetical protein
CBO3118	3315682	3316584	phosphate acetyl/butyryl transferase
CBO3120	3318277	3318747	hypothetical protein
CBO3121	3318935	3321301	recombination and DNA strand exchange inhibitor protein
CBO3123	3324374	3325252	hypothetical protein
CBO3129	3328904	3329494	hypothetical protein
CBO3130	3329699	3332080	phenylalanyl-tRNA synthetase subunit beta
CBO3131	3332241	3333260	phenylalanyl-tRNA synthetase subunit alpha
CBO3132	3333608	3334387	RNA methyltransferase, TrmH family
CBO3135	3336618	3336977	50S ribosomal protein L20
CBO3136	3337007	3337204	50S ribosomal protein L35
CBO3138	3338035	3339942	threonyl-tRNA synthetase
CBO3140	3341344	3342051	hypothetical protein
CBO3141	3342259	3343830	ABC transporter permease
CBO3143	3345434	3346324	Hsp33-like chaperonin
CBO3144	3346362	3347102	hypothetical protein
CBO3148	3349751	3350668	oligopeptide/dipeptide ABC transporter permease
CBO3149	3350684	3351613	oligopeptide/dipeptide ABC transporter permease
CBO3151	3353861	3354853	aspartate-semialdehyde dehydrogenase
CBO3152	3354876	3355754	dihydrodipicolinate synthase
CBO3156	3358699	3359409	2,3,4,5-tetrahydropyridine-2,6-carboxylate N-succinyltransferase
CBO3157	3359525	3360226	single-stranded DNA-binding protein
CBO3158	3360304	3360801	hypothetical protein
CBO3160	3361916	3362443	hypothetical protein
CBO3161	3362446	3363330	hypothetical protein
CBO3163	3363739	3365028	bifunctional folylpolyglutamate synthase/ dihydrofolate synthase
CBO3164	3365474	3368119	valyl-tRNA synthetase
CBO3171	3378617	3379378	ABC transporter ATP-binding protein
CBO3173	3380322	3380669	hypothetical protein
CBO3174	3380792	3381976	transglutaminase domain protein
CBO3176	3382949	3384178	peptidase, M16 family
CBO3177	3384399	3384902	hypothetical protein
CBO3178	3384940	3385527	flavin reductase
CBO3179	3385584	3386279	ErfK/YbiS/YcfS/YnhG family protein
CBO3180	3386314	3387189	hypothetical protein
CBO3181	3387231	3387509	hypothetical protein
CBO3182	3387637	3388443	pyrroline-5-carboxylate reductase
CBO3183	3388555	3390918	formate acetyltransferase
CBO3185	3391891	3392460	DNA-binding protein
CBO3187	3393860	3394201	hypothetical protein
CBO3188	3394213	3395949	von Willebrand factor A
CBO3189	3395972	3397300	ATPase AAA
CBO3191	3399068	3400153	hypothetical protein
CBO3192	3400332	3400526	(Fe-S)-binding protein
CBO3193	3400769	3401662	hypothetical protein
CBO3194	3401653	3402762	R-2-hydroxyglutaryl-CoA dehydratase subunit
CBO3199	3407583	3408722	butyryl-CoA dehydrogenase
CBO3200	3408916	3410109	acetyl-CoA acetyltransferase
CBO3201	3410119	3410949	3-hydroxybutyryl-CoA dehydrogenase
CBO3202	3410980	3411762	3-hydroxybutyryl-CoA dehydratase
CBO3203	3412178	3412396	hypothetical protein
CBO3204	3412440	3413537	GIY-YIG domain-containing protein
CBO3205	3413813	3415105	xanthine/uracil permease
CBO3207	3416471	3416623	hypothetical protein
CBO3209	3417009	3417773	prolipoprotein diacylglyceryl transferase
CBO3210	3417982	3418446	phosphodiesterase
CBO3212	3418985	3419740	glycerophosphoryl diester phosphodiesterase
CBO3214	3422037	3422582	rubredoxin/rubrerythrin
CBO3215	3423224	3423922	pyruvate formate-lyase activating enzyme

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO3216	3424102	3426330	formate acetyltransferase 1
CBO3217	3426842	3427093	hypothetical protein
CBO3218	3427521	3428090	ribosomal 5S rRNA E-loop binding protein Ctc/L25/TL5
CBO3220	3429097	3429525	FUR family transcriptional regulator
CBO3221	3429594	3430835	extracellular solute-binding protein
CBO3222	3430930	3431865	hypothetical protein
CBO3223	3432025	3432981	hypothetical protein
CBO3224	3433196	3433606	hypothetical protein
CBO3227	3435354	3435938	ribosome biogenesis GTP-binding protein YsxC
CBO3228	3435953	3438274	ATP-dependent protease La
CBO3230	3440219	3441508	ATP-dependent protease ATP-binding subunit ClpX
CBO3231	3441528	3442112	ATP-dependent Clp protease proteolytic subunit
CBO3232	3442238	3443530	trigger factor
CBO3234	3444600	3445412	hypothetical protein
CBO3235	3445626	3446201	orotate phosphoribosyltransferase
CBO3236	3446235	3447131	dihydroorotate dehydrogenase 1B
CBO3237	3447131	3447886	dihydroorotate dehydrogenase electron transfer subunit
CBO3239	3448906	3450096	dihydroorotase, multifunctional complex type
CBO3240	3450118	3450558	aspartate carbamoyltransferase
CBO3241	3450560	3451483	aspartate carbamoyltransferase
CBO3242	3451738	3452496	CoA-substrate-specific enzyme activase
CBO3244	3453076	3454335	R-2-hydroxyglutaryl-CoA dehydratase subunit
CBO3261	3465168	3465767	deoxyribonucleotide triphosphate pyrophosphatase
CBO3263	3467357	3467512	hypothetical protein
CBO3264	3467515	3469077	extracellular solute-binding protein, family 5
CBO3265	3469231	3470661	aspartyl/glutamyl-tRNA amidotransferase subunit B
CBO3266	3470697	3472154	aspartyl/glutamyl-tRNA amidotransferase subunit A
CBO3267	3472191	3472478	aspartyl/glutamyl-tRNA amidotransferase subunit C
CBO3268	3472613	3473824	integral membrane transport protein
CBO3269	3474121	3474390	hypothetical protein
CBO3272	3478840	3479139	TrpR family protein
CBO3273	3479437	3480153	RNA pseudouridine synthase
CBO3275	3480967	3482970	triple tyrosine motif-containing protein
CBO3276	3483104	3483271	hypothetical protein
CBO3277	3483435	3483869	hypothetical protein
CBO3280	3485795	3486439	TetR family transcriptional regulator
CBO3283	3487747	3488751	D-lactate dehydrogenase
CBO3284	3488938	3489933	D-lactate dehydrogenase
CBO3289	3495464	3496588	R-phenyllactate dehydratase small subunit
CBO3290	3496590	3497813	R-phenyllactate dehydratase medium subunit
CBO3292	3498612	3499850	E-cinnamoyl-CoA:R-phenyllactate CoA transferase large subunit
CBO3294	3502001	3503323	branched chain amino acid ABC transporter carrier protein
CBO3295	3503782	3505314	GMP synthase
CBO3296	3505327	3506781	inosine 5'-monophosphate dehydrogenase
CBO3297	3507221	3507970	PP-loop family protein
CBO3298	3508049	3509674	molecular chaperone GroEL
CBO3299	3509704	3509991	co-chaperonin GroES
CBO3303	3513799	3514716	8-oxoguanine DNA glycosylase
CBO3306	3517372	3518004	redox-sensing transcriptional repressor Rex
CBO3307	3518204	3520126	drug resistance ABC transporter ATP-binding protein
CBO3308	3520467	3521231	DNA-binding response regulator, LytTr family
CBO3309	3521232	3522866	sensor histidine kinase
CBO3310	3523109	3524527	carbon starvation protein CstA
CBO3311	3524578	3527505	peptidase
CBO3312	3527886	3531320	pyruvate carboxylase
CBO3314	3533026	3533649	endodeoxyribonuclease RusA
CBO3316	3535650	3536036	hypothetical protein
CBO3318	3537005	3537913	manganese/zinc/iron chelate uptake ABC transporter Mzt family, substrate-binding protein
CBO3320	3539490	3539948	hypothetical protein

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO3321	3539945	3540658	glycoprotease
CBO3322	3540651	3541109	ribosomal-protein-alanine acetyltransferase
CBO3323	3541467	3542690	tyrosyl-tRNA synthetase
CBO3324	3542923	3543330	hypothetical protein
CBO3325	3543466	3544623	ATP-binding protein
CBO3329	3546922	3548511	peptide chain release factor 3
CBO3330	3549016	3549822	HAD hydrolase
CBO3331	3549941	3550939	tryptophanyl-tRNA synthetase
CBO3333	3551552	3551926	desulfoferrodoxin
CBO3334	3552109	3552303	hypothetical protein
CBO3337	3562731	3563933	aspartokinase
CBO3340	3567091	3568182	hypothetical protein
CBO3342	3570597	3572573	ATP-dependent metalloprotease FtsH
CBO3343	3573000	3574391	dipeptidase PepV
CBO3348	3577282	3578940	sulfate permease
CBO3350	3581480	3582688	lipoprotein
CBO3351	3582693	3583598	bacitracin ABC transporter ATP-binding protein
CBO3355	3587039	3587932	BadF/BadG/BcrA/BcrD ATPase
CBO3356	3587992	3588288	PTS system lactose/cellobiose-specific transporter subunit IIA
CBO3357	3588530	3588838	PTS system, lactose/cellobiose-specific family, IIB component
CBO3358	3588958	3591687	sigma-54 dependent transcriptional regulator
CBO3359	3591812	3593104	PTS system, lactose/cellobiose family, IIC component
CBO3366	3599564	3599839	hypothetical protein
CBO3369	3602179	3603564	23S rRNA (uracil-5-)-methyltransferase RumA
CBO3370	3603797	3604150	hypothetical protein
CBO3371	3604477	3605487	oxidoreductase, aldo/keto reductase
CBO3372	3605631	3607388	pyruvate kinase
CBO3373	3607422	3608381	6-phosphofructokinase
CBO3375	3612343	3613290	hypothetical protein
CBO3376	3613297	3614634	hypothetical protein
CBO3377	3614631	3615515	hypothetical protein
CBO3380	3617404	3619260	excinuclease ABC subunit C
CBO3381	3619274	3619882	metal-dependent hydrolase
CBO3385	3623668	3624105	FHA domain-containing protein
CBO3386	3624204	3627026	excinuclease ABC subunit A
CBO3387	3627045	3629042	excinuclease ABC subunit B
CBO3389	3630418	3631623	carboxyl-terminal protease
CBO3390	3631833	3632723	cell division protein FtsX
CBO3391	3632713	3633399	cell division ATP-binding protein FtsE
CBO3394	3634973	3635914	transketolase, C-terminal subunit
CBO3395	3635914	3636738	transketolase
CBO3395A	3637001	3637354	PemK family protein
CBO3396	3637311	3637634	hypothetical protein
CBO3397	3637773	3638933	alanine racemase
CBO3398	3638961	3639575	hypothetical protein
CBO3399	3639775	3641277	carbohydrate kinase
CBO3400	3641274	3641654	holo-(acyl-carrier-protein) synthase
CBO3403	3644155	3645012	hypothetical protein
CBO3404	3645172	3645831	hypothetical protein
CBO3405	3646084	3646677	30S ribosomal protein S4
CBO3410	3650832	3652658	glucosamine--fructose-6-phosphate aminotransferase
CBO3411	3653261	3653938	transaldolase
CBO3412	3653957	3654604	haloacid dehalogenase
CBO3413	3654628	3655413	sorbitol-6-phosphate dehydrogenase
CBO3414	3655433	3655798	PTS system glucitol/sorbitol-specific transporter subunit IIA
CBO3415	3656014	3657021	PTS system glucitol/sorbitol-specific transporter subunit IIBC
CBO3416	3657031	3657576	PTS system glucitol/sorbitol-specific transporter subunit IIC
CBO3417	3657619	3658014	glucitol operon activator protein
CBO3418	3658028	3659077	L-iditol 2-dehydrogenase

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO3419	3659112	3660083	glucitol operon regulator
CBO3423	3663203	3664273	2-ketoisovalerate ferredoxin reductase
CBO3424	3664289	3664498	2-oxoacid:acceptor oxidoreductase subunit delta
CBO3425	3664600	3665292	hypothetical protein
CBO3426	3665525	3666595	butyrate kinase
CBO3427	3666624	3667535	phosphate butyryltransferase
CBO3428	3667528	3668622	butyrate kinase
CBO3430	3670323	3671603	hypothetical protein
CBO3432	3672676	3673209	cell wall hydrolase
CBO3433	3673601	3674464	oxidoreductase, pyridine nucleotide-disulfide family
CBO3434	3674530	3674847	thioredoxin
CBO3435	3674956	3676131	isoaspartyl dipeptidase
CBO3436	3676206	3677459	hypothetical protein
CBO3439	3680267	3682276	sensory box sigma-54 dependent transcriptional regulator
CBO3441	3683190	3683753	hydro-lyase, Fe-S type, tartrate/fumarate subfamily, beta region
CBO3442	3683845	3684687	fumarate hydratase
CBO3443	3684828	3685532	N-acetylmuramoyl-L-alanine amidase
CBO3444	3685798	3687414	Na/Pi-cotransporter family protein/PhoU family protein
CBO3445	3687720	3688112	30S ribosomal protein S9
CBO3446	3688138	3688575	50S ribosomal protein L13
CBO3447	3688696	3689436	tRNA pseudouridine synthase A
CBO3448	3689436	3690239	cobalt ABC transporter permease
CBO3449	3690236	3691102	cobalt transporter ATP-binding subunit
CBO3450	3691087	3691944	cobalt transporter ATP-binding subunit
CBO3451	3692026	3692367	50S ribosomal protein L17
CBO3452	3692411	3693358	DNA-directed RNA polymerase subunit alpha
CBO3453	3693426	3694046	30S ribosomal protein S4
CBO3454	3694081	3694479	30S ribosomal protein S11
CBO3455	3694497	3694868	30S ribosomal protein S13
CBO3457	3695111	3695329	translation initiation factor IF-1
CBO3458	3695337	3695627	hypothetical protein
CBO3459	3695633	3696382	methionine aminopeptidase
CBO3461	3697056	3698333	preprotein translocase subunit SecY
CBO3462	3698334	3698774	50S ribosomal protein L15
CBO3463	3698794	3698973	50S ribosomal protein L30
CBO3464	3698987	3699484	30S ribosomal protein S5
CBO3465	3699502	3699861	50S ribosomal protein L18
CBO3466	3699880	3700422	50S ribosomal protein L6
CBO3467	3700446	3700844	30S ribosomal protein S8
CBO3468	3700874	3701059	30S ribosomal protein S14
CBO3469	3701076	3701618	50S ribosomal protein L5
CBO3470	3701641	3701958	50S ribosomal protein L24
CBO3471	3701982	3702350	50S ribosomal protein L14
CBO3472A	3702661	3702873	50S ribosomal protein L29
CBO3474	3703324	3703995	30S ribosomal protein S3
CBO3475	3704013	3704348	50S ribosomal protein L22
CBO3476	3704370	3704654	30S ribosomal protein S19
CBO3477	3704724	3705557	50S ribosomal protein L2
CBO3478	3705612	3705905	50S ribosomal protein L23
CBO3480	3706550	3707179	50S ribosomal protein L3
CBO3481	3707269	3707577	30S ribosomal protein S10
CBO3484	3711342	3711812	30S ribosomal protein S7
CBO3485	3711973	3712350	30S ribosomal protein S12
CBO3486	3712426	3712665	ribosomal protein L7Ae family protein
CBO3487	3712790	3716326	DNA-directed RNA polymerase subunit beta'
CBO3488	3716347	3720045	DNA-directed RNA polymerase subunit beta
CBO3489	3720310	3720681	50S ribosomal protein L7/L12
CBO3490	3720728	3721228	50S ribosomal protein L10
CBO3491	3721430	3722119	50S ribosomal protein L1

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO3492	3722176	3722601	50S ribosomal protein L11
CBO3493	3722672	3723196	transcription antitermination protein NusG
CBO3494	3723243	3723470	preprotein translocase subunit SecE
CBO3495	3723520	3723669	50S ribosomal protein L33
CBO3498	3726095	3726604	hypothetical protein
CBO3499	3726607	3727380	TrmH family RNA methyltransferase
CBO3500	3727383	3728141	FAD-dependent thymidylate synthase
CBO3500A	3728125	3728565	RNase3 domain-containing protein
CBO3501	3728614	3730011	cysteinyl-tRNA synthetase
CBO3504	3733077	3733766	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
CBO3505	3733768	3734868	PIN/TRAM domain protein
CBO3506	3735273	3735671	hypothetical protein
CBO3507	3735703	3736764	DNA integrity scanning protein DisA
CBO3508	3736783	3738141	DNA repair protein RadA
CBO3510	3740698	3741741	ATP:guanido phosphotransferase
CBO3512	3742296	3742760	transciptional repressor CtsR
CBO3513	3743156	3743386	hypothetical protein
CBO3515	3743733	3745805	elongation factor G
CBO3518	3760733	3762247	lysyl-tRNA synthetase
CBO3519	3762266	3762748	transcription elongation factor GreA
CBO3521	3763756	3764727	dihydrouridine synthase
CBO3522	3764737	3765513	pantothenate kinase
CBO3525	3769431	3769976	hypoxanthine phosphoribosyltransferase
CBO3526	3769960	3771366	tRNA(Ile)-lysidine synthetase
CBO3527	3771515	3773902	stage II sporulation protein E
CBO3529	3774621	3775025	hypothetical protein
CBO3530	3775087	3775356	cell division protein FtsL
CBO3531	3775451	3775858	hypothetical protein
CBO3532	3775864	3776154	sporulation protein YabP
CBO3533	3776274	3776513	S4 domain-containing protein
CBO3534	3776587	3776865	DNA-binding protein HU
CBO3535	3776993	3778450	MazG family protein
CBO3536	3778453	3779988	stage V sporulation protein B
CBO3537	3780121	3780672	stage V sporulation protein T
CBO3538	3780871	3781881	peptidyl-prolyl isomerase
CBO3539	3781965	3785471	transcription-repair coupling factor
CBO3540	3785499	3786068	peptidyl-tRNA hydrolase
CBO3541	3786178	3787350	protease
CBO3542	3787385	3788794	sensor histidine kinase
CBO3543	3788795	3789481	DNA-binding response regulator
CBO3544	3789661	3790620	ribose-phosphate pyrophosphokinase
CBO3545	3790641	3792014	bifunctional N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate
CBO3546	3792176	3792463	regulatory protein SpoVG
CBO3547	3792566	3793381	pur operon repressor
CBO3550	3795576	3796925	[Fe] hydrogenase
CBO3551	3797011	3797655	haloacid dehalogenase
CBO3554	3799159	3800091	radical SAM family protein
CBO3556	3801435	3802481	hypothetical protein
CBO3557	3802514	3803410	radical SAM domain protein
CBO3559	3805652	3806314	DNA-binding protein
CBO3561	3806721	3807239	cyclophilin type peptidyl-prolyl cis-trans isomerase
CBO3562	3807386	3808159	glutamate racemase
CBO3563	3808352	3810250	glutamine synthetase
CBO3564	3810295	3810843	hypothetical protein
CBO3566	3811227	3812492	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
CBO3567	3812518	3813303	metallo-beta-lactamase
CBO3568	3813320	3813889	lipoprotein
CBO3569	3813904	3814407	hypothetical protein
CBO3570	3814586	3815065	rRNA large subunit methyltransferase

Core genome of *C. botulinum/C. sporogenes*

locus_tag	Minimum	Maximum	product
CBO3579	3823003	3823842	radical SAM domain protein
CBO3581	3824258	3824509	hypothetical protein
CBO3583	3825559	3826935	thiamine biosynthesis protein ThiH
CBO3584	3826944	3828143	GTP-binding protein
CBO3585	3828458	3829831	sodium:solute symporter family protein
CBO3586	3830131	3831375	glycoside hydrolase family protein
CBO3587	3831537	3832214	PHP domain-containing protein
CBO3588	3832510	3832995	hypothetical protein
CBO3589	3833102	3833533	hypothetical protein
CBO3590	3833714	3834619	LysR family transcriptional regulator
CBO3591	3834707	3835903	subtilase
CBO3592	3835938	3836819	LysR family transcriptional regulator
CBO3593	3836848	3837795	D-isomer specific 2-hydroxyacid dehydrogenase
CBO3594	3837995	3838834	acetyl-CoA carboxylase carboxyl transferase subunit alpha
CBO3595	3838881	3839750	acetyl-CoA carboxylase carboxyl transferase subunit beta
CBO3597	3841148	3841582	(3R)-hydroxymyristoyl-ACP dehydratase
CBO3598	3841602	3842078	acetyl-CoA carboxylase biotin carboxyl carrier protein
CBO3599	3842082	3843329	3-oxoacyl-ACP synthase
CBO3600	3843340	3844098	3-ketoacyl-ACP reductase
CBO3601	3844104	3845048	malonyl CoA-acyl carrier protein transacylase
CBO3602	3845069	3845995	oxidoreductase, 2-nitropropane dioxygenase
CBO3603	3846121	3846345	acyl carrier protein
CBO3604	3846386	3847366	3-oxoacyl-ACP synthase
CBO3604A	3847359	3847835	MarR family transcriptional regulator
CBO3605	3847851	3848909	oxidoreductase, 2-nitropropane dioxygenase
CBO3606	3849006	3849941	peptidase
CBO3608	3851101	3852078	DNA replication protein DnaC
CBO3611	3853324	3854133	hypothetical protein
CBO3612	3854146	3854520	CoA-binding protein
CBO3613	3854701	3855930	pyridine nucleotide-disulfide oxidoreductase
CBO3615	3856884	3857069	hypothetical protein
CBO3616	3857195	3858481	adenylosuccinate synthetase
CBO3617	3858713	3860098	pyridine nucleotide-disulfide oxidoreductase
CBO3618	3860264	3860479	hypothetical protein
CBO3619	3860672	3862393	phosphoglucomutase/phosphomannomutase
CBO3620	3862635	3863972	replicative DNA helicase
CBO3621	3864003	3865922	ATP-dependent protease La
CBO3622	3865933	3866376	50S ribosomal protein L9
CBO3624	3868362	3869366	hypothetical protein
CBO3625	3869378	3869695	hypothetical protein
CBO3626	3869843	3870085	30S ribosomal protein S18
CBO3628	3870566	3870850	30S ribosomal protein S6
CBO3630	3872378	3872575	hypothetical protein
CBO3631	3872591	3873460	mechanosensitive ion channel family protein
CBO3632	3873467	3873712	hypothetical protein
CBO3633	3873903	3875057	cysteine desulfurase
CBO3635	3876184	3876765	hypothetical protein
CBO3636	3876767	3877012	hypothetical protein
CBO3637	3877065	3877574	hypothetical protein
CBO3638	3877612	3878460	stage 0 sporulation protein J
CBO3639	3878476	3879240	sporulation initiation inhibitor protein soj
CBO3640	3879367	3880146	parB family protein
CBO3641	3880246	3880965	16S rRNA methyltransferase GidB
CBO3642	3880989	3882866	tRNA uridine 5-carboxymethylaminomethyl modification protein GidA
CBO3643	3882875	3884260	tRNA modification GTPase TrmE
CBO3644	3884308	3884946	DNA/RNA-binding protein

Appendix D. Effect of the addition of *C. botulinum* and *C. sporogenes* genomes (from left to right) on the number of CDS represented in *'n'* genomes. Similarity in CDS was calculated at an 80% identity cut-off using BLASTN.

	Clostridium botulinum str. Hall	*Clostridium botulinum* A str. ATCC 19397	*Clostridium botulinum* A str. ATCC 3502	*Clostridium botulinum* F str. 230613	*Clostridium botulinum* F str. Langeland	*Clostridium botulinum* Ba4 str. 657	*Clostridium botulinum* H04402 065	*Clostridium botulinum* A2 str. Kyoto	*Clostridium botulinum* A3 str. Loch Maree	*Clostridium botulinum* B1 str. Okra	*Clostridium sporogenes* str. ATCC 15579	*Clostridium sporogenes* str. PA 3679
n=1	3368	203	470	1107	598	1118	1580	1756	2109	2286	3048	3645
n=2		3336	284	220	707	663	641	719	722	731	794	1056
n=3			3124	714	452	411	384	412	421	428	451	486
n=4				2548	359	238	239	200	201	237	247	260
n=5					2535	513	201	189	179	152	157	163
n=6						2316	602	217	171	154	149	143
n=7							2165	579	217	141	126	112
n=8								2126	598	217	159	149
n=9									2034	596	272	198
n=10										2001	723	386
n=11											1726	611
n=12												1590
TOTAL CDS	**3368**	**3539**	**3878**	**4589**	**4651**	**5259**	**5812**	**6198**	**6652**	**6943**	**7852**	**8799**

Appendix E. Identified signature genes in *C. sporogenes* NCTC 8594 and ATCC 15579 from pan-genomic analysis

locus_tag	Length	Product	COG Group
IYC_00747	243	hypothetical protein	
IYC_00947	4,293	putative CoA-substrate-specific enzyme activase	COG1924 Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)
IYC_00987	249	hypothetical protein	
IYC_01319	207	hypothetical protein	
IYC_01464	>193	hypothetical protein	
IYC_01469	873	NAD-dependent epimerase/dehydratase	COG0451 Nucleoside-diphosphate-sugar epimerases; overlaps another CDS with the same product name
IYC_01474	963	NAD-dependent epimerase/dehydratase	COG0451 Nucleoside-diphosphate-sugar epimerases; overlaps another CDS with the same product name
IYC_01479	2,058	glycosyl transferase family protein	COG0438 Glycosyltransferase
IYC_01669	1,017	glyceraldehyde-3-phosphate dehydrogenase, type I	COG0057 Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase
IYC_01904	1,707	putative methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_02034	738	hypothetical protein	
IYC_02184	591	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_02189	699	phosphonate-transporting ATPase	COG1136 ABC-type antimicrobial peptide transport system, ATPase component
IYC_02194	3,105	hypothetical protein	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_02249	2,472	efflux ABC transporter permease	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_02259	1,014	two-component system histidine kinase	COG0642 Signal transduction histidine kinase
IYC_02289	456	acetyltransferase	COG2153 Predicted acyltransferase
IYC_02349	144	hypothetical protein	
IYC_02434	627	deoxyuridine 5'-triphosphate nucleotidohydrolase	COG0756 dUTPase
IYC_02439	192	small acid-soluble spore protein, H-type	
IYC_02554	2,232	restriction endonuclease family protein	COG1002 Type II restriction enzyme, methylase subunits
IYC_02629	579	hypothetical protein	
IYC_02694	1,062	permease	COG4767 Glycopeptide antibiotics resistance protein
IYC_02709	687	NPQTN cell wall surface anchor protein	COG5386 Cell surface protein
IYC_02714	894	iron (III) dicitrate-binding periplasmic protein	COG0614 ABC-type Fe3+-hydroxamate transport system,periplasmic component
IYC_02724	777	ferric enterobactin transport ATP-binding protein fepC	COG1120 ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components
IYC_02729	609	sortase	COG3764 Sortase (surface protein transpeptidase)
IYC_02769	1,092	putative permease	COG0701 Predicted permeases
IYC_02889	786	AP endonuclease	COG1082 Sugar phosphate isomerases/epimerases
IYC_03014	366	hypothetical protein	
IYC_03034	897	HDIG domain-containing protein	COG3481 Predicted HD-superfamily hydrolase
IYC_03496	516	hypothetical protein	
IYC_03511	276	hypothetical protein	
IYC_03536	933	phospholipase family protein	COG1752 Predicted esterase of the alpha-beta hydrolase superfamily
IYC_03586	618	thiamine-phosphate pyrophosphorylase	COG0352 Thiamine monophosphate synthase
IYC_03681	231	hypothetical protein	COG3655 Predicted transcriptional regulator
IYC_03701	>617	hypothetical protein	
IYC_03716	792	endonuclease/exonuclease/phosphatase family protein	
IYC_03751	990	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_03791	582	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_03796	1,044	hypothetical protein	
IYC_03843	987	putative transporter	COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily
IYC_03848	369	hypothetical protein	COG2832 Uncharacterized protein conserved in bacteria
IYC_03853	1,806	ABC transporter, permease/ATP-binding protein	COG4988 ABC-type transport system involved in cytochrome bd biosynthesis, ATPase and permease components
IYC_03878	915	iron chelate ABC transporter solute-binding protein	COG0614 ABC-type Fe3+-hydroxamate transport system,periplasmic component
IYC_03888	762	iron chelate ABC transporter ATP-binding protein	COG1120 ABC-type cobalamin/Fe3+-siderophores transport systems, ATPase components
IYC_03898	849	lipoyl synthase	COG0320 Lipoate synthase
IYC_03913	846	RpiR family transcriptional regulator	COG1737 Transcriptional regulators
IYC_03933	852	AraC family transcriptional regulator	COG2207 AraC-type DNA-binding domain-containing proteins
IYC_03938	420	hypothetical protein	
IYC_03943	921	radical SAM-superfamily protein	COG1533 DNA repair photolyase
IYC_03988	543	hypothetical protein	
IYC_03998	1,185	peptidase	COG1473 Metal-dependent amidase/aminoacylase/carboxypeptidase
IYC_04008	648	amino acid ABC transporter permease	COG0765 ABC-type amino acid transport system, permease component
IYC_04078	1,767	helicase domain-containing protein	COG0513 Superfamily II DNA and RNA helicases
IYC_04228	639	putative methionine-S-sulfoxide reductase	COG0225 Peptide methionine sulfoxide reductase
IYC_04278	1,035	hypothetical protein	
IYC_04313	624	MutT/NUDIX NTP pyrophosphatase	COG0494 NTP pyrophosphohydrolases including oxidative damage repair enzymes
IYC_04343	810	polar amino acid ABC transporter amino acid-binding protein	COG0834 ABC-type amino acid transport/signal transduction systems, periplasmic component/domain
IYC_04523	1,452	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase	COG0769 UDP-N-acetylmuramyl tripeptide synthase
IYC_04528	1,389	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	COG0770 UDP-N-acetylmuramyl pentapeptide synthase
IYC_04628	537	bifunctional pyrimidine regulatory protein PyrR uracil phosphoribosyltransferase	COG2065 Pyrimidine operon attenuation protein/uracil phosphoribosyltransferase
IYC_04783	2,214	efflux ABC transporter, permease protein	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_04788	624	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_04983	675	two component transcriptional regulator, winged helix family protein	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_04988	1,035	sensory transduction histidine kinase	COG0642 Signal transduction histidine kinase
IYC_05003	1,905	hypothetical protein	COG0577 ABC-type antimicrobial peptide transport system, permease component
IYC_05293	1,638	putative prolyl-tRNA synthetase	COG0442 Prolyl-tRNA synthetase

IYC_05579	1,524	Mg chelatase-like protein	COG0606 Predicted ATPase with chaperone activity
IYC_05724	696	Clp protease	COG0740 Protease subunit of ATP-dependent Clp proteases
IYC_05814	606	hypothetical protein	
IYC_05819	810	hypothetical protein	COG1484 DNA replication protein
IYC_05884	927	hypothetical protein	
IYC_06029	630	regulatory protein TetR	COG1309 Transcriptional regulator
IYC_06044	1,377	ABC transporter	COG1122 ABC-type cobalt transport system, ATPase component
IYC_06049	2,115	type I restriction-modification system specificity subunit	COG0286 Type I restriction-modification system methyltransferase subunit
IYC_06054	1,380	putative type I restriction system, specificity protein HsdS	COG0732 Restriction endonuclease S subunits
IYC_06089	2,187	hypothetical protein	COG0451 Nucleoside-diphosphate-sugar epimerases
IYC_06094	888	hypothetical protein	
IYC_06099	1,830	hypothetical protein	COG4878 Uncharacterized protein conserved in bacteria
IYC_06104	1,416	hypothetical protein	COG0438 Glycosyltransferase
IYC_06109	1,470	hypothetical protein	COG4267 Predicted membrane protein
IYC_06114	1,698	spore coat protein	
IYC_06119	702	hypothetical protein	
IYC_06124	729	hypothetical protein	
IYC_06204	1,791	methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_06349	210	ferrous iron transport protein	COG1918 Fe2+ transport system protein A
IYC_06369	198	hypothetical protein	COG2323 Predicted membrane protein
IYC_06374	438	acetyltransferase	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_06494	507	hypothetical protein	COG0452 Phosphopantothenoylcysteine synthetase/decarboxylase
IYC_06499	327	hypothetical protein	
IYC_06514	1,794	TPR repeat-containing glycosyl transferase	COG0463 Glycosyltransferases involved in cell wall biogenesis
IYC_06586	1,116	putative FAD-dependent dehydrogenase	COG0644 Dehydrogenases (flavoproteins)
IYC_06681	8,631	cyclic beta 1-2 glucan synthetase	COG3459 Cellobiose phosphorylase
IYC_06781	204	hypothetical protein	
IYC_06998	1,356	MATE efflux family protein	COG0534 Na+-driven multidrug efflux pump
IYC_07148	1,083	streptolysin associated protein SagC	COG0476 Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2
IYC_07178	180	hypothetical protein	
IYC_07285	945	DNA polymerase III subunit delta'	COG2812 DNA polymerase III, gamma/tau subunits
IYC_07325	642	hypothetical protein	
IYC_07685	852	N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase	COG2890 Methylase of polypeptide chain release factors
IYC_07815	1,515	F0F1 ATP synthase subunit alpha	COG0056 F0F1-type ATP synthase, alpha subunit
IYC_07825	1,389	F0F1 ATP synthase subunit beta	COG0055 F0F1-type ATP synthase, beta subunit
IYC_07935	1,650	hypothetical protein	
IYC_08020	1,659	putative threonyl-tRNA synthetase/uridine kinase	COG0441 Threonyl-tRNA synthetase
IYC_08050	1,941	hypothetical protein	COG1032 Fe-S oxidoreductase
IYC_08638	819	hypothetical protein	
IYC_08643	627	hypothetical protein	
IYC_08713	1,494	[Fe] hydrogenase	COG1145 Ferredoxin
IYC_08949	897	hypothetical protein	
IYC_09089	1,029	hypothetical protein	
IYC_09204	342	hypothetical protein	COG0222 Ribosomal protein L7/L12
IYC_09209	462	hypothetical protein	
IYC_09234	456	hypothetical protein	COG0454 Histone acetyltransferase HPA2 and related acetyltransferases
IYC_09509	405	hypothetical protein	
IYC_09514	345	hypothetical protein	
IYC_09519	378	ErfK/YbiS/YcfS/YnhG family protein	COG1376 Uncharacterized protein conserved in bacteria
IYC_09529	681	hypothetical protein	COG3861 Uncharacterized protein conserved in bacteria
IYC_09654	186	hypothetical protein	
IYC_09659	207	hypothetical protein	
IYC_09774	195	hypothetical protein	COG1476 Predicted transcriptional regulators
IYC_09779	219	hypothetical protein	
IYC_09794	1,335	MerR family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_09854	1,524	ferrous iron transport protein B	COG0370 Fe2+ transport system protein B
IYC_09969	294	hypothetical protein	
IYC_10119	201	DNA-binding protein	COG1396 Predicted transcriptional regulators
IYC_10134	1,152	hypothetical protein	COG5279 Uncharacterized protein involved in cytokinesis, contains TGc (transglutaminase/protease-like) domain
IYC_10139	1,911	cell wall binding repeat 2-containing protein	COG2247 Putative cell wall-binding domain
IYC_10159	762	beta-lactamase domain-containing protein	COG0491 Zn-dependent hydrolases, including glyoxylases
IYC_10164	162	hypothetical protein	
IYC_10284	558	hypothetical protein	
IYC_10359	609	hypothetical protein	COG4832 Uncharacterized conserved protein
IYC_10364	1,077	hypothetical protein	
IYC_10369	816	MerR family transcriptional regulator	COG0789 Predicted transcriptional regulators
IYC_10669	690	hypothetical protein	
IYC_10709	825	metallo-beta-lactamase family protein	COG1237 Metal-dependent hydrolases of the beta-lactamase superfamily II
IYC_10829	684	DNA-binding response regulator	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_10934	285	YCII-like protein	COG2350 Uncharacterized protein conserved in bacteria
IYC_10939	594	transcriptional regulator	COG1309 Transcriptional regulator
IYC_11064	1,842	putative spermidine/putrescine ABC transporter protein	COG1177 ABC-type spermidine/putrescine transport system, permease component II; putative spermidine/putrescine ABC transporter, permease protein/spermidine/putrescine-binding protein
IYC_11079	1,698	putative methyl-accepting chemotaxis protein	COG0840 Methyl-accepting chemotaxis protein
IYC_11334	177	transposase	COG3666 Transposase and inactivated derivatives
IYC_11389	2,679	putative diguanylate cyclase	COG2199 FOG: GGDEF domain
IYC_11509	468	hypothetical protein	

IYC_11559	696	lrgB family protein	COG1346 Putative effector of murein hydrolase
IYC_11769	1,497	methyl-accepting chemotaxis sensory transducer	COG0840 Methyl-accepting chemotaxis protein
IYC_11944	552	hypothetical protein	
IYC_11949	3,234	helicase, SNF2/RAD54 family protein	COG0553 Superfamily II DNA/RNA helicases, SNF2 family
IYC_11974	1,596	peptide/opine/nickel uptake ABC transporter substrate-binding protein	COG0747 ABC-type dipeptide transport system, periplasmic component
IYC_12259	375	S-adenosylmethionine decarboxylase proenzyme	COG1586 S-adenosylmethionine decarboxylase
IYC_12394	1,005	hydrogenase expression/formation protein HypE	COG0309 Hydrogenase maturation factor
IYC_12464	459	transcription activator effector binding protein	COG4978 Transcriptional regulator, effector-bindingdomain/component
IYC_12499	1,188	monovalent cation:proton antiporter-2 (CPA2) family protein	COG0025 NhaP-type Na+/H+ and K+/H+ antiporters
IYC_12554	1,059	endonuclease/exonuclease/phosphatase family protein	
IYC_12649	198	thiamine biosynthesis protein ThiS	COG2104 Sulfur transfer protein involved in thiamine biosynthesis
IYC_12664	1,158	thiamine biosynthesis protein ThiH	COG1060 Thiamine biosynthesis enzyme ThiH and related uncharacterized enzymes
IYC_12904	429	hypothetical protein	
IYC_12949	1,656	hypothetical protein	COG3949 Uncharacterized membrane protein
IYC_13014	1,488	integral membrane sensor signal transduction histidine kinase	COG0642 Signal transduction histidine kinase
IYC_13019	693	response regulator receiver	COG0745 Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain
IYC_13024	459	ubiquitin-associated protein	COG1308 Transcription factor homologous to NACalpha-BTF3
IYC_13089	1,731	peptidoglycan binding protein	COG4932 Predicted outer membrane protein
IYC_13094	672	peptidase, sortase like protein	COG3764 Sortase (surface protein transpeptidase)
IYC_13134	747	putative acetyltransferase	COG0110 Acetyltransferase (isoleucine patch superfamily)
IYC_13199	1,083	putative oxidoreductase	COG0644 Dehydrogenases (flavoproteins)
IYC_13424	555	sporulation protein YunB	
IYC_13479	549	nitroreductase	COG0778 Nitroreductase
IYC_13739	1,833	phosphoglycerol transferase	COG1368 Phosphoglycerol transferase and related proteins, alkaline phosphatase superfamily
IYC_13744	237	hypothetical protein	
IYC_13749	1,407	glutamate decarboxylase	COG0076 Glutamate decarboxylase and related PLP-dependent proteins
IYC_13978	264	PRC-barrel domain-containing protein	COG1873 Uncharacterized conserved protein
IYC_14103	1,218	M16 family peptidase	COG0612 Predicted Zn-dependent peptidases
IYC_14133	378	hypothetical protein	
IYC_14203	687	VanZ family membrane protein	COG4767 Glycopeptide antibiotics resistance protein
IYC_14293	471	hypothetical protein	
IYC_14348	1,176	GAF sensor signal transduction histidine kinase	COG0642 Signal transduction histidine kinase
IYC_14368	405	putative chloride channel protein EriC	
IYC_14393	420	MarR family transcriptional regulator	COG1846 Transcriptional regulators
IYC_14453	894	N-acetylmuramic acid-6-phosphate etherase	COG2103 Predicted sugar phosphate isomerase
IYC_14463	858	BglG family transcriptional antiterminator	COG3711 Transcriptional antiterminator
IYC_14478	198	hypothetical protein	COG1476 Predicted transcriptional regulators
IYC_14603	2,646	magnesium-translocating P-type ATPase	COG0474 Cation transport ATPase
IYC_14688	1,197	aspartate/ornithine carbamoyltransferase family protein	COG0078 Ornithine carbamoyltransferase
IYC_14708	1,764	sigma-54 dependent transcriptional regulator	COG3829 Transcriptional regulator containing PAS, AAA-type ATPase, and DNA-binding domains
IYC_14763	546	hypothetical protein	COG1434 Uncharacterized conserved protein
IYC_15068	1,089	hypothetical protein	COG2391 Predicted transporter component
IYC_15073	1,158	cysteine desulfurase family protein	COG0520 Selenocysteine lyase
IYC_15613	1,158	flagellar hook-associated protein 3	COG1344 Flagellin and related hook-associated proteins
IYC_15808	177	hypothetical protein	
IYC_15833	669	capsular polysaccharide biosynthsis protein	COG2148 Sugar transferases involved in lipopolysaccharide synthesis
IYC_15903	828	hypothetical protein	COG2755 Lysophospholipase L1 and related esterases
IYC_15988	1,311	flagellar protein export ATPase FliI	COG1157 Flagellar biosynthesis/type III secretory pathway ATPase
IYC_16118	2,547	putative calcium-translocating P-type ATPase, PMCA-type	COG0474 Cation transport ATPase
IYC_16143	327	putative V-type ATPase, G subunit	COG3599 Cell division initiation protein
IYC_16148	1,959	V-type ATP synthase subunit I	COG1269 Archaeal/vacuolar-type H+-ATPase subunit I
IYC_16153	489	V-type ATP synthase subunit K	COG0636 F0F1-type ATP synthase, subunit c/Archaeal/vacuolar-type H+-ATPase, subunit K
IYC_16158	600	V-type ATPase, E subunit	COG1390 Archaeal/vacuolar-type H+-ATPase subunit E
IYC_16163	1,002	V-type ATP synthase subunit C	COG1527 Archaeal/vacuolar-type H+-ATPase subunit C
IYC_16198	291	hypothetical protein	
IYC_16293	462	hypothetical protein	COG0613 Predicted metal-dependent phosphoesterases (PHP family)
IYC_16473	1,032	hypothetical protein	COG1559 Predicted periplasmic solute-binding protein
IYC_16758	1,665	STE like transcription factor domain-containing protein	COG5298 Uncharacterized protein conserved in bacteria
IYC_16763	1,263	group 2 family glycosyl transferase	COG1215 Glycosyltransferases, probably involved in cell wall biogenesis
IYC_16813	2,373	cation-transporting ATPase, P-type	COG0474 Cation transport ATPase
IYC_16823	795	CAAX amino terminal protease family protein	COG1266 Predicted metal-dependent membrane protease
IYC_16878	429	hypothetical protein	
IYC_17018	654	putative inner membrane protein translocase component YidC	COG0706 Preprotein translocase subunit YidC
IYC_17098	1,149	2-hydroxyglutaryl-CoA dehydratase subunit D	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB
IYC_17103	1,149	2-hydroxyglutaryl-CoA dehydratase subunit D	COG1775 Benzoyl-CoA reductase/2-hydroxyglutaryl-CoAdehydratase subunit, BcrC/BadD/HgdB
IYC_17113	447	single-strand binding protein	COG0629 Single-stranded DNA-binding protein
IYC_17513	471	D-proline reductase proprotein PrdA	
IYC_17773	828	hypothetical protein	
IYC_17840	1,584	hypothetical protein	
IYC_17875	1,956	putative tetracycline resistance protein	COG0480 Translation elongation factors (GTPases)

IYC_18220	927	hypothetical protein	COG2188 Transcriptional regulators
IYC_18225	354	PRD domain-containing protein	COG3711 Transcriptional antiterminator
IYC_18230	360	hypothetical protein	
IYC_18235	1,299	hypothetical protein	
IYC_18240	873	aryldialkylphosphatase	COG1735 Predicted metal-dependent hydrolase with the TIM-barrel fold
IYC_18245	1,110	hypothetical protein	COG0520 Selenocysteine lyase
IYC_18250	1,161	putative amino acid racemase	COG3457 Predicted amino acid racemase
IYC_18255	1,209	putative mutase	COG1015 Phosphopentomutase
IYC_18260	726	glucosamine-6-phosphate deaminase	COG0363 6-phosphogluconolactonase/Glucosamine-6-phosphate isomerase/deaminase
IYC_18480	936	hypothetical protein	COG0697 Permeases of the drug/metabolite transporter (DMT) superfamily
IYC_18490	666	chloramphenicol acetyltransferase	COG4845 Chloramphenicol O-acetyltransferase
IYC_18630	759	hypothetical protein	COG0550 Topoisomerase IA
IYC_18650	630	putative chloramphenicol acetyltransferase	COG0110 Acetyltransferase (isoleucine patch superfamily)
IYC_18710	756	TPR repeat-containing protein	COG0457 FOG: TPR repeat
IYC_18730	282	hypothetical protein	
IYC_18735	327	hypothetical protein	
IYC_18740	366	hypothetical protein	
IYC_18875	330	hypothetical protein	COG2801 Transposase and inactivated derivatives
IYC_18890	612	TetR family transcriptional regulator	COG1309 Transcriptional regulator
IYC_18895	2,568	pyruvate phosphate dikinase PEP/pyruvate-bindingprotein	COG0574 Phosphoenolpyruvate synthase/pyruvate phosphate dikinase
IYC_19100	990	hypothetical protein	
IYC_19110	1,089	hypothetical protein	
IYC_19150	822	diaminopimelate epimerase	COG0253 Diaminopimelate epimerase
IYC_19175	453	hypothetical protein	
IYC_19260	597	stage III sporulation protein AG	
IYC_19360	885	ferredoxin-NADP(+) reductase subunit alpha	COG0543 2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases
IYC_19385	831	iron-sulfur binding protein	COG1149 MinD superfamily P-loop ATPase containing an inserted ferredoxin domain
IYC_19460	699	hypothetical protein	COG2234 Predicted aminopeptidases
IYC_19480	969	aliphatic sulfonate ABC transporter substrate-binding protein	COG0715 ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components
IYC_19485	711	aliphatic sulfonates ABC transporter ATP-bindingprotein	COG1116 ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component
IYC_19495	612	hypothetical protein	COG4624 Iron only hydrogenase large subunit, C-terminal domain
IYC_19690	249	hypothetical protein	
IYC_19695	147	hypothetical protein	
IYC_19715	246	hypothetical protein	
IYC_19775	1,038	Peptidoglycan-binding lysin domain protein	COG1388 FOG: LysM repeat
IYC_19850	531	hypothetical protein	COG0558 Phosphatidylglycerophosphate synthase
IYC_19940	360	hypothetical protein	COG4728 Uncharacterized protein conserved in bacteria
IYC_19955	1,392	hypothetical protein	
IYC_20190	297	sporulation protein YqfC	
IYC_20290	2,595	putative cation transporting ATPase	COG0474 Cation transport ATPase
IYC_20295	1,398	alginate O-acetyltransferase	
IYC_20300	1,416	alginate O-acetylation protein algI	COG1696 Predicted membrane protein involved in D-alanine export
IYC_20460	585	Maf-like protein	COG0424 Nucleotide-binding protein implicated in inhibition of septum formation
IYC_20555	348	N-acetylmuramoyl-L-alanine amidase	COG3103 SH3 domain protein; overlaps another CDS with the same product name
IYC_20565	>185	hypothetical protein	
IYC_20731	1,563	putative sensory box sigma-54 dependent transcriptional regulator	COG3283 Transcriptional regulator of aromatic aminoacids metabolism
IYC_20741	447	cytidine/deoxycytidylate deaminase family protein	COG0590 Cytosine/adenosine deaminases

Appendix F. Identified signature genes in *C. botulinum* strains from pan-genomic analysis

locus_tag	Length	Product
CBO0030	438	cytidine/deoxycytidylate deaminase
CBO0056	945	DNA polymerase III subunit delta'
CBO0064	612	hypothetical protein
CBO0138	852	protein methyltransferase
CBO0152	480	ATP synthase F0F1 subunit B
CBO0154	1515	ATP synthase F0F1 subunit alpha
CBO0156	1,392	ATP synthase F0F1 subunit beta
CBO0158	708	hypothetical protein
CBO0179	1,650	hypothetical protein
CBO0202	1938	hypothetical protein
CBO0223	726	polysaccharide deacetylase
CBO0252	513	GNAT family acetyltransferase
CBO0257	675	isochorismate family protein
CBO0305	972	AraC family transcriptional regulator
CBO0335	453	hypothetical protein
CBO0439	933	phospholipase
CBO0449	618	thiamine-phosphate pyrophosphorylase
CBO0469	240	hypothetical protein
CBO0488	1083	streptolysin associated protein SagC
CBO0544	462	GNAT family acetyltransferase
CBO0617	1032	hypothetical protein
CBO0677	753	hypothetical protein
CBO0734	201	DNA-binding protein
CBO0836	513	RNA polymerase sigma-70 factor, ECF family
CBO0837	1164	hypothetical protein
CBO0840	813	non-heme chloroperoxidase
CBO0858	516	hypothetical protein
CBO0876	501	peptide deformylase
CBO0879	807	(Fe-S)-binding protein
CBO0898	378	ErfK/YbiS/YcfS/YnhG family protein
CBO0899	330	hypothetical protein
CBO1027	759	hypothetical protein
CBO1032	438	acetyltransferase
CBO1074	381	hypothetical protein
CBO1095	1017	glyceraldehyde-3-phosphate dehydrogenase, type I
CBO1102	438	MarR family transcriptional regulator
CBO1176	741	hypothetical protein
CBO1198	891	AraC family transcriptional regulator
CBO1203	1629	beta-amylase
CBO1235	627	deoxyuridine 5'-triphosphate nucleotidohydrolase
CBO1236	192	small acid-soluble spore protein, H-type
CBO1248	222	DNA-binding protein
CBO1250	267	hypothetical protein
CBO1266	2226	restriction endonuclease
CBO1279	366	hypothetical protein
CBO1296	1653	STE like transcription factor domain protein
CBO1297	1263	glycosyl transferase family protein
CBO1337	915	iron chelate uptake ABC transporter FeCT family,solute-binding protein
CBO1339	762	iron ABC transporter ATP-binding protein
CBO1352	543	hypothetical protein
CBO1381	939	ABC transporter ATP-binding protein
CBO1391	639	methionine-S-sulfoxide reductase
CBO1405	624	pyrophosphatase, MutT/nudix family
CBO1411	807	polar amino acid ABC transporter amino acid-binding protein
CBO1457	1452	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase
CBO1458	1,389	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D- alanine ligase
CBO1478	537	bifunctional pyrimidine regulatory protein PyrR/uracil phosphoribosyltransferase
CBO1539	1404	transporter
CBO1608	336	hypothetical protein
CBO1652	162	hypothetical protein
CBO1668	552	hypothetical protein
CBO1670	3,234	helicase, SNF2/RAD54 family
CBO1769	960	kinase, PfkB family
CBO1850	1188	transporter monovalent cation:proton antiporter-2 (CPA2) family
CBO1870	615	hypothetical protein
CBO1889	597	stage III sporulation protein AG
CBO1906	453	hypothetical protein
CBO1912	822	diaminopimelate epimerase
CBO1915	258	hypothetical protein
CBO1919	1710	methyl-accepting chemotaxis protein
CBO1973	1068	alanine racemase
CBO1975	1503	spore germination protein
CBO1978	1110	spore germination protein
CBO1983	756	hypothetical protein
CBO2001	114	hypothetical protein
CBO2004	1,065	oxidoreductase
CBO2021	666	chloramphenicol acetyltransferase
CBO2037	648	hypothetical protein
CBO2038	2,478	thrombospondin type 3 repeat protein
CBO2051	1,062	glycosyl transferase family protein
CBO2052	645	transferase, hexapeptide repeat family
CBO2053	1,140	DegT/DnrJ/EryC1/StrS family aminotransferase
CBO2054	900	glycosyl transferase family protein
CBO2055	1062	glucose-1-phosphate thymidylyltransferase
CBO2056	981	dTDP-glucose 4,6-dehydratase
CBO2101	924	UDP-glucose epimerase
CBO2130	789	hypothetical protein
CBO2132	489	DNA binding protein, excisionase
CBO2193	1632	hypothetical protein
CBO2230	342	hypothetical protein
CBO2231	1,956	oligopeptide transporter OPT family
CBO2240	450	GNAT family acetyltransferase

CBO2241	465	MarR family transcriptional regulator
CBO2242	405	hypothetical protein
CBO2270	582	hypothetical protein
CBO2272	204	hypothetical protein
CBO2291	822	transglutaminase
CBO2292	996	hypothetical protein
CBO2306	1251	sensor histidine kinase
CBO2307	681	DNA-binding response regulator
CBO2409	693	Clp protease
CBO2531	264	PRC-barrel domain-containing protein
CBO2548	285	hypothetical protein
CBO2550	804	recombinase, phage RecT family
CBO2604	840	hypothetical protein
CBO2626	1002	V-type ATP synthase subunit C
CBO2627	603	V-type ATPase, E subunit
CBO2628	477	V-type ATP synthase subunit K
CBO2634	2547	calcium-translocating P-type ATPase
CBO2659	1,317	flagellar protein export ATPase FliI
CBO2678	1431	polysaccharide biosynthesis protein
CBO2679	918	UDP-glucose 4-epimerase
CBO2681	1662	cytidylyltransferase domain-containing protein
CBO2682	1170	hypothetical protein
CBO2685	1101	UDP-galactopyranose mutase
CBO2687	711	capsular exopolysaccharide family protein
CBO2691	177	hypothetical protein
CBO2739	963	flagellar hook-associated protein 3
CBO2766	1209	peptidase, M16 family
CBO2771	387	hypothetical protein
CBO2804	426	hypothetical protein
CBO2836	858	BglG family transcriptional antiterminator
CBO2847	717	CoA-transferase subunit A
CBO2850	489	dehydrogenase, [2Fe-2S] binding subunit
CBO2851	831	dehydrogenase, FAD binding subunit
CBO2852	2,331	dehydrogenase, molybdenum binding subunit
CBO2853	1,218	transporter monovalent cation:proton antiporter-2 (CPA2) family
CBO2869	2646	magnesium-translocating P-type ATPase
CBO2885	1197	aspartate/ornithine carbamoyltransferase
CBO2907	1632	peptidase
CBO2908	498	hypothetical protein
CBO2951A	297	sporulation protein YqfC
CBO2971	2595	cation transporting ATPase
CBO2972	1515	hypothetical protein
CBO3004	585	Maf-like protein
CBO3047	1452	hypothetical protein
CBO3082	555	hypothetical protein
CBO3097	1128	glycoside hydrolase family protein
CBO3102	1,104	hypothetical protein
CBO3103	1,419	oligosaccharide repeat-containing polymerase
CBO3104	1095	capsular polysaccharide biosynthesis protein
CBO3105	1,308	polysaccharide biosynthesis protein
CBO3106	747	acetyltransferase
CBO3111	2904	N-acetylmuramoyl-L-alanine amidase
CBO3125	153	hypothetical protein
CBO3162	261	hypothetical protein
CBO3206	897	HDIG domain protein
CBO3211	366	hypothetical protein
CBO3225	429	hypothetical protein
CBO3233	786	AP endonuclease, family 2
CBO3286	1197	electron transfer flavoprotein subunit alpha
CBO3302	1,497	[Fe] hydrogenase
CBO3338	1,299	aminopeptidase
CBO3349	1386	sensor histidine kinase
CBO3352	204	hypothetical protein
CBO3353	996	hypothetical protein
CBO3360	1,338	glycosyl hydrolase family protein
CBO3503	1,638	prolyl-tRNA synthetase
CBO3509	2436	negative regulator of genetic competence MecB/ClpC
CBO3552	399	hypothetical protein
CBO3582	1,047	biotin synthase
CBO3627	447	single-strand binding protein
CBO3629	1149	2-hydroxyacyl-CoA dehydratase subunit beta
CBO3645	654	inner membrane protein translocase component YidC